

Bidrag til overlevelsesmodellene i livsforsikring

av

Lars Østhassel

MASTEROPPGAVE

for graden

Master i Modellering og Dataanalyse

(Master of Science)



Det matematisk- naturvitenskapelige fakultet
Universitetet i Oslo

Mai 2013

Faculty of Mathematics and Natural Sciences
University of Oslo

Annerkjennelser

Først og fremst ønsker jeg å rekke en stor takk til min veileder Erik Bølviken. For det første har han gitt meg en spennende og interessant oppgave. Jeg har hatt stor glede av de mange samtaler rundt oppgaven og aktuarfaget generelt. Bølviken har også vært min foreleser i en rekke aktuarkurs på Universitet. Pensumet har vært særdeles nyttig i denne oppgaven. I tillegg har Bølviken bistått med programkode.

Tusen takk til Finn for gode innvendinger og takk til Rebecca for hjelp med asymptotisk teori. Videre ønsker jeg å takke mine medstudenter på lesesal B802 for et godt miljø.

Jeg vil også takke min familie for en utrolig støtte gjennom min tid på Universitet. Spesielt Ivar som har holdt motivasjonen min oppe samtidig bidratt med interessante diskusjoner. Til slutt vil jeg takke min kjæreste Silje som har vært en uvurdelig støtte.

Sammendrag

I denne oppgaven brukes Kullback-Leibler divergens og bootstrap-simuleringer til å studere feil i velkjente dødelighetsmodeller anvendt i livsforsikring. Feil deles inn i tre kategorier; estimering-, modell- og seleksjonsfeil.

Modellfeilen i Gompertz-Makeham modellen er avtagende i perioden 1960 til 2009. Dessuten er det større avvik i russiske menn sammenlignet med en rekke andre land. Estimeringsfeilen holder seg konstant i tilsvarende periode.

Innhold

1	Innledning	1
I	Bakgrunn	3
2	Dødelighetsmodellering	5
2.1	Statisk dødelighet	5
2.1.1	Standard Gompertz-Makeham	6
2.1.2	Dødelighet med ekstra usikkerhet	7
2.2	Dynamisk dødelighet	9
2.2.1	Lee-Carter modellen	11
3	Feilanalyse i dødelighetsmodellering ved pensjonsberegninger	13
3.1	Teoretisk feilanalyse I	13
3.1.1	Dødelighet med bootstrap og parameterfeil	13
3.1.2	Dødelighet med asymptotisk matematikk	15
3.2	Teoretisk feilanalyse II	17
3.2.1	Kullback-Leibler distansen	17
3.2.2	Feil i antatt modell	18
3.3	Pensjon- og dødelighetsfeil i lengre levetid	20
4	Datasett og Estimeringsmetoder	21
4.1	Datasett	21
4.2	Gompertz-Makeham	23
4.3	Beta-binomisk med dødelighet	27
4.4	Lee-Carter	28
II	Numeriske resultater	33
5	Gompertz-Makeham	35
5.1	Avvik i parametre og modell	35
5.1.1	År 2009	35
5.1.2	Periode 1960 til 2008	40
5.2	Avvik i pensjonsevalueringer	44

5.2.1	År 2009	44
5.2.2	Periode 1960 til 2008	52
5.3	Asymptotikk mot bootstrapping	55
5.4	Oppsummering	56
6	Lee-Carter	59
6.1	Sensitivitet	59
6.2	Backtesting	61
6.3	Pensjonsevalueringer med lengre levetid	61
6.4	Oppsummering	63
7	Konklusjoner	65
7.1	Bemerkninger	65
7.2	Veien videre	65
III	Vedlegg	67
8	Vedlegg	69
8.1	Vedlegg A - Formler	69
8.2	Vedlegg B - Programkode	71
8.2.1	Kapittel 2	71
8.2.2	Kapittel 4	71
8.2.3	Kapittel 5	73
8.2.4	Kapittel 6	87
	Bibliografi	94

Kapittel 1

Innledning

Usikkerhet i livsforsikring fører til feil i premier, avsetninger og forpliktelser, og deles inn i to kjernepunkter, demografisk og finansiell feil. I denne oppgaven studeres utelukkende det førstnevnte slik at avkastning og inflasjon antas å være gitt i markedet. Utfordringen i demografiske feil er i stor grad tilknyttet feil i *dødelighetsmodellering*, og målet er å beskrive den fremtidige dødelighetsutviklingen så presist som mulig gjennom en *dødelighetsmodell*.

Siden 1800-tallet har det vært mange bidrag og utvidelser innenfor dødelighetsmodellering. I 1825 introduserte Benjamin Gompertz, *Gompertz law of mortality* som senere ble videreutviklet av William Makeham. *Gompertz-Makeham* modellen har i stor grad preget dødelighetsmodellering, og brukes fremdeles idag i mange ulike former og varianter. Feil i dødelighet skyldes at mennesker lever enten for kort eller for lenge i henhold til den estimerte levealderen. Hovedproblemet idag for livsforsikringsselskaper er lengre levetid.

Helt siden andre verdenskrig har dødeligheten i de industrielle landene gått kraftig ned, spesielt for barn og unge. De fremtidige prognosene regner med at for hvert tiår vil den gjennomsnittlige levealderen øke med ett år. Det nye dødelighetsmønsteret har gitt behov for å introdusere *dynamiske* representasjoner som er en utvidelse av de *statiske* dødelighetsmodellene. Den mest aktuelle idag, *Lee-Carter* modellen, ble publisert i 1992 av Ronald Lee og Lawrance Carter. Modellen har i de senere årene vokst i popularitet, og er idag den viktigste modellen til å predikere lengre levetid.

Det er knyttet tre utfordringer til feil i modellering av lengre levetid: Estimerings-, modell- og seleksjonsfeil. Estimeringsfeilen i denne oppgaven avhenger av *eksponeringsstørrelsen* på pensjonsporteføljer, og er sensitiv for små porteføljer. Modellfeil er derimot uavhengig av både historisk dødelighetsdata og eksponeringsstørrelse hvor feil i modellen systematisk avviker virkeligheten. Det betyr at modellen aldri gjenspeiler virkeligheten eksakt. Seleksjonsproblemet er å ta utgangspunkt i riktig populasjon og dessuten være oppmerksom på foreldet dødelighetsdata ettersom dødelighetene er i konstant utvikling. Det er store avvik innad i en befolkning hvor dødelighetsnivået til en pensjonsportefølje ligger langt lavere enn landsgjennomsnit-

tet. Slike faktorer reduserer selvsagt mengden av aktuell dødelighetsdata og eksponeringsstørrelse. Det motiverer for gruppeinndeling ved pensjonsberegninger.

I denne oppgaven anvendes teori som gjør det mulig å analysere de tre feilene ovenfor anvendt på engangspremier. Målet er se hvordan feilene påvirker populasjonsgrupper både internasjonalt og nasjonalt ved Norge, Russland, Italia, Sverige, Japan og USA. For å kunne gjennomføre en slik feilanalyse er det nødvendig å kjenne til den korrekte dødeligheten. Tanken er å bruke kontrollerte eksponeringsstørrelser på gitte modeller samtidig som korrekt dødelighet allerede er kjent. Det vil gi et sett med henholdsvis korrekt og parametrisert engangspremier som gjør det mulig å analysere estimering og modellfeil for en gitt livsforsikringssituasjon.

Utgangspunktet og ideen som ligger til grunn er å anta et korrekt sett med dødeligheter. Det tillater oss å bruke *Kullback-Leibler* divergensen som determinerer det parametriserte settet som ligger nærmest det korrekte settet med dødeligheter. Estimeringen viser seg å være uavhengig av eksponeringsstørrelse slik at avviket mellom settene er en direkte årsak for at modellen ikke gjenspeiler virkeligheten eksakt, altså modellfeil. En sentral del av oppgaven er å simulere dødeligheter ved bootstrap for diverse eksponeringsstørrelser. Ved tilstrekkelig mengde eksponering vil bootstrap-estimatene konvergere mot Kullback-Leibler estimatet slik at total gjestående feil skyldes feil i modellen.

Det betyr at feilene mellom bootstrap-simuleringene og Kullback-Leibler settet skyldes estimeringsfeil. Dette oppstår når eksponeringsstørrelsen ikke er tilstrekkelig stor. Det gir oss muligheten til å analysere hvor godt den underliggende modellen beskriver det korrekte settet med dødeligheter, og hvor stor eksponeringsstørrelse i en pensjonsportefølje som er nødvendig for at modellfeilen dominerer. Dette gjøres som nevnt for en rekke land som gjør det mulig å analysere den siste årsaken til feil i livsforsikring, altså seleksjonsfeil.

Kapitlene 2 og 3 er teoretisk bakgrunn med henholdsvis dødelighetsmodellering og feilanalyse. Datasett, estimeringsmetoder og pseudo-koder er plassert i kapittel 4. Kapitlene 5 og 6 inneholder numeriske resultater for henholdsvis Gompertz-Makeham og Lee-Carter modellen.

Del I

Bakgrunn

Kapittel 2

Dødelighetsmodellering

Utfordringen i dødelighetsmodellering er å finne den riktige dødelighet i alder x til $x+1$ definert ved g_x . Siden fremtiden er ukjent eksisterer ikke denne dødeligheten. Istedenfor brukes historisk dødelighetsdata i tråd med statistiske dødelighetsmodeller til å finne disse dødelighetene så nøyaktig som mulig, uttrykt ved q_x . Videre deles dødelighetsmodellene inn i to grupper, statiske og dynamiske modeller. I seksjon 2.1 introduseres førstnevnte sammen med Gompertz-Makeham modellen. I seksjon 2.2 defineres dynamisk dødelighet og Lee-Carter modellen.

2.1 Statisk dødelighet

En statisk dødelighetsrepresentasjon tar utgangspunkt at dødeligheten i neste periode er lik dødeligheten idag. Det tas ikke hensyn til fremtidige dødelighetstrender og baserer seg kun på nåtidens dødelighetsnivå. Det betyr at en statisk dødelighetsmodell beregner dødelighet kun for en periode.

Dødelighetstabell

Anta at q_x og overlevelsessannsynlighet $p_x = 1 - q_x$ er kjent. En *dødelighetstabell*[19] n_x , er definert ved antall gjenlevende i alder x slik at

$$n_{x+1} = n_x \cdot p_x \text{ for } x = 0, 1, \dots, x_e - 1 \quad (2.1)$$

hvor x_e er antatt maksimal alder i en befolkning. Det er en oversikt som viser, for hver alder x , hva sannsynligheten for å dø innen neste alder $x+1$. En tilsvarende metode er å bruke en dødelighetsfrekvens[19] μ_x hvor antall gjenlevende er gitt ved en overlevelsesfunksjon

$$n_x = n_0 \cdot e^{-\int_0^{x_e} \mu_t dt}$$

hvor n_0 er antall levende det første året.

Det er vanlig å bruke dødelighetstabeller i forbindelse med dødelighetsmodellering. På nettsiden human mortality database[23] er det publisert dødelighetstabeller for en rekke land. Det er ofte et godt utgangspunkt å måle diverse usikkerheter tilknyttet dødelighet.

Aldersprofil og pensjonsporteføljer

En aldersprofil beskriver formen til en populasjon, for eksempel en undergruppe i en befolkning. I livsforsikring er pensjonsporteføljer ofte en slik undergruppe. Pensjonsporteføljer eksponeres for risiko i alle aldre hvor majoriteten er yrkesaktive i alderen 30-65 år. En enkel måte er å definere en aldersprofil[1] følgende

$$w_x = c \exp(-\gamma|x - \bar{x}|) \text{ der } \sum_x w_x = 1$$

hvor \bar{x} er alderen med flest levende og γ beskriver formen. For en gitt eksponeringsstørrelse N blir *pensjonsporteføljen*¹ i denne oppgaven ordnet slik at $N_x = N \cdot w_x$. Aldersprofilen viser seg å være særdeles nyttig når overlevelsesmodellene blir introdusert.

En overlevelsesmodell har som mål å beskrive fremtidig aldersmønster på grunnlag av historisk dødelighetsdata. Nedenfor er den mest utbredte definert, Gompertz-Makeham modellen. Videre er det gjort noen utvidelser.

2.1.1 Standard Gompertz-Makeham

I 1825 foreslo Benjamin Gompertz[10] at dødeligheten vokser eksponentialt, og definerte følgende funksjon, Gompertz Law of mortality

$$\mu_x = b \cdot e^{cx}$$

hvor b er grunnlinjen og c er aldringskomponenten. William Makeham[18] videreutviklet modellen ved å legge til en konstant a slik at intensiteten er følgende

$$\mu_x = a + b \cdot e^{cx}$$

Makeham bidraget er uavhengig av alder, mens Gompertz leddet er avhengig. Gompertz-Makeham er en av de mest utbredte[19] dødelighetsmodellene i verden, og derfor et naturlig valg. Sannsynligheten for å overleve fra alder x til $x+1$ er gitt[1]

$$\begin{aligned} p_{\theta x} &= \exp(-\mu_x) \\ &= \exp(-a - b \cdot e^{cx}) \text{ hvor } \theta = (\alpha, \beta, c) > 0 \end{aligned}$$

¹La $N = 1\,000\,000$ med aldersprofil der $\gamma = 0.10$ og $\bar{x} = 40$. Det vil gi en pensjonsportefølje med overkant av 800 000 mennesker i alderen 30 til 65 år. Tanken er at hvert menneske i pensjonsprofilen N_x kjøper en engangspremie i alder x . Eksponeringsstørrelsen N omtales da både som datagrunnlaget i selve estimeringen av dødeligheten, og størrelsen på en pensjonsportefølje. En pensjonsportefølje med en million mennesker vil omlag 40 000 være 40 år med aldersprofilen ovenfor.

2.1. STATISK DØDELIGHET

slik at Gompertz-Makeham dødeligheten er gitt[1]

$$q_{\theta x} = 1 - \exp(-a - b \cdot e^{c \cdot x})$$

som er dødeligheten i alder x til $x + 1$. Sannsynligheten for å overleve fra x til $x+k$ år er gitt ved integrasjonen.

$$\begin{aligned} {}_k p_{\theta x} &= \exp\left(\int_x^{x+k} -\mu_x dy\right) \\ &= \exp\left(-a \cdot x - \frac{b}{c} \cdot e^{c \cdot x}\right) \end{aligned}$$

Dersom $q_{\theta x}$ er kjent, er $p_{\theta x} = 1 - q_{\theta x}$. Sannsynligheten for å dø k perioder frem i tid finner vi gjennom rekursjonen[1]

$${}_{k+1} q_{\theta x} = q_{\theta, x+k} \cdot {}_k p_{\theta x}, \quad k = 0, 1, \dots$$

hvor ${}_0 p_l = 1$.

Gompertz-Makeham som en parametrisk funksjon

For en gitt, stor populasjon N_x , der hvert dødsfall opptrer uavhengig av hverandre. En naturlig antakelse er at antall døde y_x er binomisk fordelt med aldersprofil N_x og dødelighet q_x . La $q_{\theta x}$ være Gompertz-Makeham dødeligheten slik at den parametriske funksjonen anvendt med binomisk teori ovenfor blir

$$f_{\theta x} = \prod_{i=1}^x \binom{N_x}{y_x} q_{\theta x}^{y_x} (1 - q_{\theta x})^{N_x - y_x}$$

Dette resultatet skal brukes senere i forbindelse med å estimere parameterene (a,b,c).

2.1.2 Dødelighet med ekstra usikkerhet

Standard Gompertz-Makeham modellen beskriver ikke usikkerhet direkte. Nedenfor er beta-binomisk fordeling med en underliggende Gompertz-Makeham modell blitt brukt til å beskrive ekstra usikkerhet i dødelighet.

Beta-fordeling

Beta-fordelingen[2] er en familie av kontinuerlige sannsynlighetsfordelinger definert i intervallet (0,1) og er vanligvis beskrevet av to ikke-negative parametre α og β . Beta-fordelingen med en tilfeldig variabel x i intervallet (0,1) er definert følgende

$$g(x) = \frac{1}{\beta(\alpha, \beta)} \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1} \quad \text{hvor} \quad \beta(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

der Γ er en vanlig gammafunksjon. Videre er forventning og standardavvik gitt

$$E(q) = \frac{\alpha}{\alpha + \beta}$$

$$Sd(q) = \sqrt{\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}}$$

Beta-binomisk fordeling med Gompertz-Makeham dødelighet

Beta-fordelingen blir ofte brukt som en *konjugat prior* for binomisk fordeling. Modellen er godt egnet til å beregne utvikling av tilfeldige variable i intervallet (0,1). Ideen videre er å dra nytte av dette. Bruker egenskapen om at y_x er binomisk fordelt slik at likelihooden er gitt

$$\Pr(Y = y_x | q) = \binom{n_x}{y_x} \cdot q^{y_x} \cdot (1 - q)^{n_x - y_x}$$

med en konjugat beta-fordeling gitt

$$\pi(q | \beta, \alpha) = \frac{q^{\alpha-1}(1-q)^{\beta-1}}{\beta(\alpha, \beta)}$$

Bruker nå egenskapen til en *sammensatt sannsynlighetsfordeling* (se vedlegg) slik at sannsynligheten for at antall y dør i alder x er gitt

$$\begin{aligned} \Pr(Y = y) &= \int \Pr(Y = y | q) \pi(q | \alpha, \beta) dq \\ &= \int_0^1 \binom{n}{y} \cdot q^y \cdot (1 - q)^{n-y} \cdot \frac{q^{\alpha-1}(1-q)^{\beta-1}}{\beta(\alpha, \beta)} dq \\ &= \binom{n}{y} \frac{\beta(y + \alpha, n - y + \beta)}{\beta(\alpha, \beta)} \\ &= \binom{n}{y} \frac{\Gamma(y + \alpha) \Gamma(n - y + \beta) \Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + n) \Gamma(\alpha) \Gamma(\beta)} \end{aligned}$$

som er en beta-binomisk[11] sannsynlighetsmodell. Det betyr at antall døde y_x er beta-binomisk fordelt med forventning og standardavvik

$$E(y_x) = \frac{N_x \alpha}{\alpha + \beta}$$

$$Sd(y_x) = \sqrt{\frac{N_x \alpha \beta (\alpha + \beta + N_x)}{(\alpha + \beta)^2 (\alpha + \beta + 1)}}$$

Ideen videre er å bygge en modell med underliggende Gompertz-Makeham dødeligheter. Modellen blir impementert gjennom en transformasjon slik at den forventede verdien til beta-binomiske modellen er lik Gompertz-Makeham dødeligheten $\xi = 1 - \exp(-a - be^c)$ og $\alpha + \beta = \gamma$. Det gir $\frac{\alpha}{\alpha + \beta} = \xi \Leftrightarrow \alpha = \gamma\xi, \beta = \gamma(1 - \xi)$ slik at dødeligheten i alder x kan nå uttrykkes ved

$$q_{\theta x}^e = \text{Beta}(\gamma \cdot \xi, \gamma \cdot (1 - \xi)) \text{ hvor } \xi = (a, b, c)$$

2.2. DYNAMISK DØDELIGHET

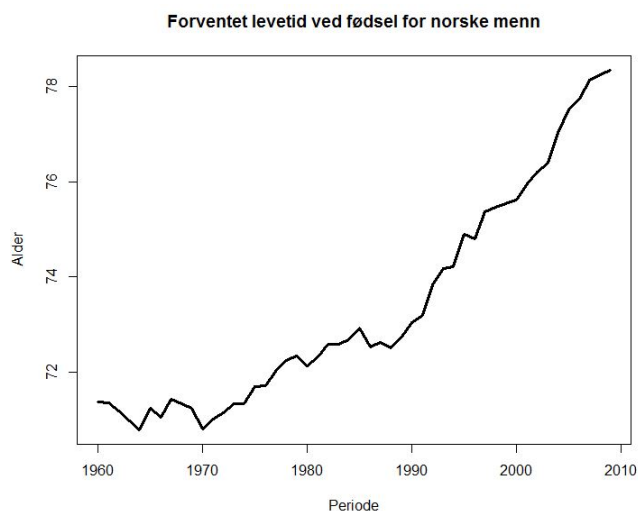
transformasjonen fører til en økning fra 2 til 4 parametre, og modellen gir en mulighet til å beregne en ekstra usikkerhet. Det kan vises matematisk når $\gamma \rightarrow \infty$ blir modellen en standard Gompertz-Makeham. En annen måte å forklare dette er å ta utgangspunkt i standardavviket til modellen

$$\sigma_{q_{\theta x}^e} = \sqrt{\frac{\xi(1-\xi)}{\gamma+1}}$$

som viser at $\sigma_{q_{\theta x}^e} \rightarrow 0$ når $\gamma \rightarrow \infty$ slik at modellen blir en standard Gompertz-Makeham.

2.2 Dynamisk dødelighet

Problemet med modellene ovenfor er den statiske representasjonen. Dødelighet predikert i lengre perioder er ikke lett å beregne, men en statisk dødelighetsmodell er ikke løsningen. Det er statistisk dokumentert[19] at dødelighetstrendene i det 20. århundre har gått kraftig ned slik at nye *dynamiske* bidrag har vært nødvendig. I livsforsikring er det viktig å skille mellom pensjon- og dødelighetsforsikring. I tider med dødelighetsreduksjon vil den statiske representasjonen ved dødelighetsforsikring tilfalle en fordel for forsikringsgiver. Det motsatte er tilfelle ved beregning av pensjonsforsikring. I denne oppgaves studeres det sistnevnte.



Figur 2.1: Figur viser utviklingen til forventet levealder på norske menn i perioden 1940-2009. Dette illustrerer hvor aktuelt det er med dynamiske modeller.

Lengre levetid har helt klart påført livselskapene kraftige kostnader og kutt. De økonomiske konsekvensene av lengre levetid vil senere bli illustrert i kapittel 6. Figur

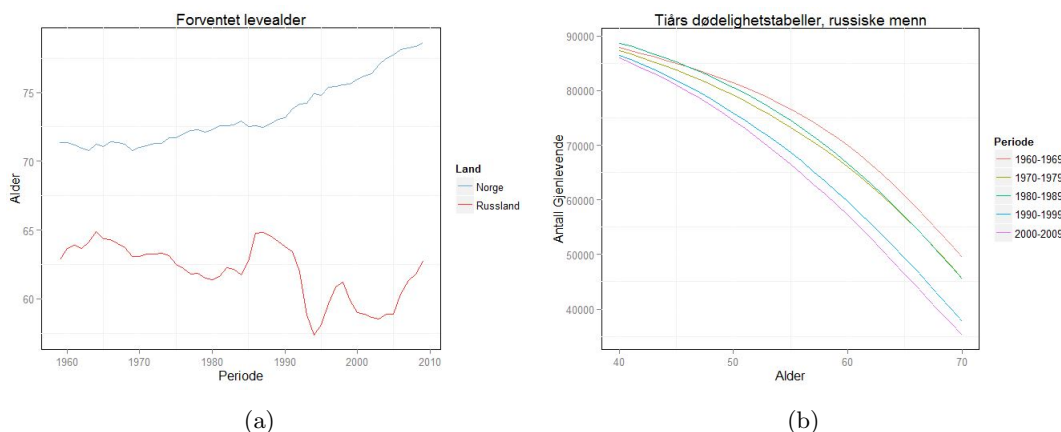
2.2. DYNAMISK DØDELIGHET

2.1 viser forventet levetid ved fødsel for norske menn i perioden 1940-2009, og illustrerer kraftig vekst. Det nye dødelighetsmønsteret krever nye metoder til å predikere fremtidig dødelighet. Dette er et kjent problem i det akademiske aktuarmiljøet og flere *dynamiske dødelighetsmodeller* er publisert. På generelt grunnlag er en dynamisk modell en utvidelse av en statisk modell hvor modellfunksjonen[19] nå avhenger av to komponenter, alder x og kalenderår k . Kalenderåret er knyttet til fremtiden slik at $q_{x,k}$ representerer modellert dødelighet i alder x til $x+1$, k år frem i tid. Den mest aktuelle modellen idag er *Lee-Carter modellen*[17], som blir introdusert i neste del-seksjon.

Et interessant spørsmål er hvor lenge mennesker lever. Det er knyttet stor usikkerhet ved å predikere dødelighet i høy alder. Kan det være grunnlag til å hevde at flere mennesker blir 110 år siden flere overlever aldrene 60-70 år. I en befolkning er det få som når denne alderen slik at de historiske dataene er små, som igjen fører til usikkerhet. Når levealderen befinner seg på et høyt nivå, blir det stadig vanskeligere å øke den enda mer. Det er to grunner til dette[14]. Det ene er at alle mennesker skal dø, og det er en biologisk oppfatning at flere vil dø av alderdom. Den andre grunnen er knyttet til de økonomiske faktorene ved lengre levetid. Forholdet mellom antall pensjonister og yrkesaktive jevnes ut, hvor samfunnet ikke har økonomi til å behandle alle på lik linje.

Russiske menn - et unntak

Russland er et interessant land hvor dødeligheten på *russiske menn* økte kraftig på 1990-tallet. Figur 2.2 a) sammenligner forventet levealder mellom norske og russiske menn fra 1960 til 2009 hvor det er tydelig forskjell. Høyre graf viser tiårs dødelighetstabeller i tilsvarende periode i alderen 40-70 år. Det er flere dødsfall per år ettersom tiden går. En sentral del av oppgaven er å sammenligne norske og russisk menn ved forskjellige dødelighetsmodeller.



Figur 2.2: Figur a) sammenligner forventet levealder mellom norske og russiske menn i perioden 1959-2009. Figur b) viser historiske livtabeller de siste 50 årene. Merk økningen i antall dødsfall.

Et studie[25] gjort i 2009 påviste at 52 % av alle dødsfall i aldergruppen 15-54 år

skyldes alkoholrelaterte problemer. Det har ført til en negativ dødelighetstrend siden første verdenskrig. I andre europeiske land utgjorde dette kun 4 %. Fra midten av 1990 har det blitt gjennomført anti-alkohol kampanjer, og figuren ovenfor viser tegn til bedring. Det kan være krevende å modellere dødelighet til et land med så usikker og variert utvikling. Lee-Carter modellen i neste del-seksjon tar utgangspunkt at dødeligheten synker. Det kan derfor gi unøyaktige resultater på russisk data uten modifikasjoner.

2.2.1 Lee-Carter modellen

Modellen ble publisert av demografen/økonomen Ronald Lee og sosionomen Lawrence Carter i 1992[17], og har som hovedmål å predikere fremtidig dødelighet for en gitt befolkning. En viktig forutsetning for modellen er at dødelighetsnivået i fremtiden fortsetter å synke. Motivasjonen er å bruke historisk dødelighetsdata. Lee og Carter publiserte modellen med utgangspunkt i amerikanske dødelighetsdata fra perioden 1933 til 1987. Resultatet[17] viste at den forventede levalder i 2065 vil stige med 10.5 år til 86 år². Idag lever 46 % av befolkningen til de er 80 år, innen 2065 vil 46 % leve til de er 90 år. Dette var langt større estimer enn det *Social Security Administration's Office of the Actuary* estimerte, som i ettertid har valgt å bruke deler av modellen selv[16].

Matematisk, er modellen definert følgende[17]

$$\log(m_{x,k}) = a_x + b_x k_k + e_{x,k} \quad (2.2)$$

hvor $m_{x,k}$ er historiske dødeligheter, a_x beskriver den gjennomsnittlige logaritmen til $m_{x,k}$, b_x beskriver avviket mellom dødeligheten i forskjellige aldre når parameteren k_k varierer. Sistnevnte er en stokastisk tidsrekke og beskriver dødelighetstrend. Feilleddet $e_{x,k}$ er normalfordelt med forventning 0 og varians σ^2 som i utgangspunktet skal ha liten betydning. Grunnen til det er at variansen er forklart gjennom tidsrekken k .

Det finnes ingen entydig løsning av ligning 2.2. Lee og Carter introduserte betingelsene $\sum_x b_x = 1$ og $\sum_k k_k = 0$ for å oppfylle det kritiske kravet om entydig løsning. Leseren kan da enkelt bekrefte at a_x faktisk er gjennomsnittet av $\log(m_{x,k})$. Lineær regresjon er ikke egnet siden høyre side av ligningen ikke har noen observerte verdier. Fremgangsmåten er å bruke *singulærverdidekomposisjonen* som løser *minste kvadraters metode* som vil gi estimer for a_x og b_x . Deretter predikere dødelighetstrenden k_k ved hjelp av stokastiske tidsrekker. Til slutt ender opp med dødeligheter

$$\hat{m}_{x,k} = \exp \left(\hat{a}_x + \hat{b}_x \hat{k}_k \right)$$

²Gjelder både menn og kvinner

2.2. DYNAMISK DØDELIGHET

som er predikert k år frem i tid for aldrene x .

Lee-Carter modellen ble primært tilpasset og anvendt for amerikanske dødelighetsdata. Idag er det utført en rekke studier og utvidelser av modellen slik at den kan anvendes for andre land. I denne oppgaven skal brukes en modifisert Lee-Carter modell som ble publisert i 2003 av Brouhn, Denuit og Vermunt[3].

Lee-carter modell med poisson fordeling

Ideen bak modellen er basert på antakelsen om at feilledet i normalfordelingen ovenfor er urealistisk. Anta at antall døende $D_{x,k}$ i alder x og periode k er poisson fordelt. Modellen er da definert følgende[3]

$$D_{x,k} \sim \text{Poisson}(E_{x,k} \cdot \mu_{x,k}) \text{ hvor } \mu_{x,k} = \exp(a_x + b_x k_k)$$

hvor parametrene a_x , b_x og k_k har tilsvarende funksjon som i originalversjonen. Istedenfor singularverdi dekomposisjonen estimeres modellen gjennom maximum likelihood estimator. Det er en fordel som skal dras nytte av i kapittel 4.

Kapittel 3

Feilanalyse i dødelighetsmodellering ved pensjonsberegninger

En sentral del av oppgaven er å analysere modell og estimeringsfeil i pensjonsporteføljer. I dette kapittelet redegjøres teori som gjør det mulig å analysere feilene separat. Hensikten er å vise at estimeringsfeilen går mot 0 når eksponeringen på en pensjonsportefølje er tilstrekkelig stor. Feilen som gjenstår skyldes dødelighetsmodellen som systematisk avviker virkeligheten.

3.1 Teoretisk feilanalyse I

3.1.1 Dødelighet med bootstrap og parameterfeil

La $\theta = (\theta_1, \dots, \theta_w)$ være parameteren som brukes for å finne den parametriske dødelighetsmodellen $f_{\theta x}$ som måler dødeligheten $\underline{q}_{\theta} = (q_{\theta 1}, q_{\theta 2}, \dots, q_{\theta x_e})$ og en ønsket interesse $\psi_{\theta} = \psi(\underline{q}_{\theta})$ i livsforsikring, for eksempel prisen på en pensjonspremie. Det optimale er å finne korrekt dødelighet $\underline{g} = (g_1, g_2, \dots, g_{x_e})$ og $\psi = \psi(\underline{g})$, men det er ikke mulig siden \underline{g} ligger utenfor parameterrommet til f. Istedenfor vil målet være å få avstanden $\psi - \psi_{\theta}$ så liten som mulig. ψ_{θ} determineres alltid gjennom estimering, og for enkelthetsens skyld vil denne refereres til $\psi_{\hat{\theta}}$ videre i oppgaven. Det finnes en rekke utmerkede estimeringsmetoder hvor *maximum likelihood estimator* ofte er foretrukket.

Det er viktig å analysere feil i parameterene. Det er ofte stor variasjon mellom parametrene som vil påvirke sluttresultatet $\psi_{\hat{\theta}}$. Det er derfor viktig å analysere og finne de parametrene med størst betydning for å redusere avviket $\psi - \psi_{\hat{\theta}}$ mest mulig. *Bootstrap*[20] er et viktig verktøy i dødelighetsmodellering, spesielt for å analysere parameterfeil. Tanken er å bruke en korrekt dødelighetstabell $Z = n_x$ som determinerer \underline{g} . Sistnevnte brukes til å trekke m_b datasett \hat{Z}^* fra en aldersprofil N_x med

3.1. TEORETISK FEILANALYSE I

eksponeringsstørrelse N . Trekningen baserer seg på antakelsen at antall døde y_x er binomisk fordelt. Datasettene blir deretter brukt til å estimere $m_b \hat{\theta}^*$ ved en parameterisk dødelighetsmodell

$$f_{\theta x} = \prod_{i=1}^x \binom{n_x}{y_x} q_{\theta x}^{y_x} (1 - q_{\theta x})^{n_x - y_x}$$

ved hjelp av maximum likelihood estimator. Videre gir det $m_b \underline{q}_{\hat{\theta}^*} = (q_{\hat{\theta}^*1}, q_{\hat{\theta}^*2}, \dots, q_{\hat{\theta}^*x_e})$ og $\psi_{\hat{\theta}}^* = \psi(\underline{q}_{\hat{\theta}^*})$. Bootstrap-prosessen illustreres følgende

$$Z \longrightarrow \underline{g} \longrightarrow m_b \hat{Z}^* \longrightarrow m_b \hat{\theta}^* \longrightarrow m_b \underline{q}_{\hat{\theta}^*} \longrightarrow m_b \psi_{\hat{\theta}^*}$$

Forventningen og standardavviket til bootstrapestimatet er definert ved[20]

$$E(\hat{\theta}^*) = \frac{1}{m_b} \sum_{i=1}^{m_b} \hat{\theta}_i^* \quad \text{og} \quad Sd(\hat{\theta}^*) = \sqrt{\frac{1}{m_b} \sum_{i=1}^{m_b} (\hat{\theta}_i^* - \bar{\hat{\theta}}^*)^2}$$

Hovedmålet og det som faktisk er interessant er å beregne interessen ψ nøyaktig som mulig. Det er derfor naturlig å se på både forventning og standardavvik til $\psi_{\hat{\theta}^*}$

$$E(\psi_{\hat{\theta}^*}) = \frac{1}{m_b} \sum_{i=1}^{m_b} \psi_{\hat{\theta}^*i} \quad \text{og} \quad Sd(\psi_{\hat{\theta}^*}) = \sqrt{\frac{1}{m_b} \sum_{i=1}^{m_b} (\psi_{\hat{\theta}^*i} - \bar{\psi}_{\hat{\theta}^*})^2}$$

Total feil mellom bootstrap-simuleringene og korrekt interesse deles opp følgende

$$|\psi - \psi_{\hat{\theta}}^*| = |\psi - \psi_{\hat{\theta}} + \psi_{\hat{\theta}} - \psi_{\hat{\theta}}^*|$$

hvor $\psi - \psi_{\hat{\theta}}$ og $\psi_{\hat{\theta}} - \psi_{\hat{\theta}}^*$ er forårsaket av to komponenter, henholdsvis estimering/modell og simulering. Sistnevnte i oppgaven avhenger av bootstrap. Antall bootstrap er valgfritt siden dagens datamaskiner kan håndtere store mengder data på få sekunder. Det antas derfor videre at simuleringskomponenten bortfaller[1]. Når simuleringskomponenten bortfaller gjenstår feilen mellom ψ og $\psi_{\hat{\theta}}^*$, deriblant forårsaket av *parameterfeil*.

Metoder for å studere parameterfeil

Det skal primært brukes to metoder, *forholdstest* og *parameterskjevhets*. En forholdstest måler forventning i forhold til standardavvik, uttrykt matematisk[1]

$$\hat{\theta}_{Forhold} = \frac{E(\hat{\theta})}{Sd(\hat{\theta})} \quad \text{og} \quad \hat{\psi}_{Forhold} = \frac{E(\hat{\psi})}{Sd(\hat{\psi})}$$

¹Videre i oppgaven vil det referes kun til $\hat{\theta}$ og $\psi_{\hat{\theta}}$ der det skrives om bootstrap.

3.1. TEORETISK FEILANALYSE I

En høy verdi indikerer lav usikkerhet. Parameterskjevheten er den relative avstanden mellom bootstrap og en gitt parameter θ_0^2 , uttrykt matematisk[20]

$$\hat{\theta}_{\text{Skjevhet}} = \frac{E(\hat{\theta} - \theta_0)}{\theta_0} \quad \text{og} \quad \hat{\psi}_{\text{Skjevhet}} = \frac{E(\psi_{\hat{\theta}} - \psi_0)}{\psi_0}$$

og er egnet til å vurdere parameterene opp mot hverandre. θ_0 introduseres i neste seksjon, og forklarer hvorfor den brukes til å måle skjevheten til bootstrap estimatorene.

3.1.2 Dødelighet med asymptotisk matematikk

I forrige seksjon ble total feil representert ved

$$|\psi - \psi_{\hat{\theta}}^*| = | \underbrace{\psi - \psi_{\hat{\theta}}}_{\text{Estimering/Modell}} + \underbrace{\psi_{\hat{\theta}} - \psi_{\hat{\theta}}^*}_{\text{Simulering}} |$$

hvor simuleringskomponenten bortfaller slik at estimering og modellkomponenten gjenstår. I utgangspunktet er den korrekte dødeligheten \underline{g} ukjent, og dessuten ligger utenfor parameterrommet til f. En ide er å introdusere en vilkårlig θ_0 som determinerer dødeligheten $\underline{q}_{\theta_0} = (q_{\theta_0 1}, q_{\theta_0 2}, \dots, q_{\theta_0 x_e})$ og $\psi_0 = \psi(\underline{q}_{\theta_0})$ slik at gjenstående feil dekomponeres ved

$$|\psi - \psi_{\hat{\theta}}| = |\psi - \psi_0 + \psi_0 - \psi_{\hat{\theta}}|$$

I teorien fra *Kullback-Leibler divergens*[15] vises det i neste seksjon at θ_0 er parametersettet som minimerer avviket og gir minst mulig feil mellom korrekt og parametrisert interesse. Forskjellen nå, er $\psi_{\hat{\theta}}$ avhenger av den begrensede eksponeringsstørrelsen N_x i en pensjonsportefølje, mens ψ_0 er konstant og dermed uavhengig av eksponeringen. Mer om dette nedenfor og i neste seksjon.

Henviser leseren til Huber[12] og White[21] for resultatene som følger. La $\hat{\theta}$ og θ_0 henholdsvis være mle og den optimale parameteren for θ . Det kritiske er å vise at $\psi_0 - \hat{\psi} \rightarrow 0$ når eksponeringsstørrelsen $N \rightarrow \infty$. Følgende resultater må holde

1. $\theta_0 - \hat{\theta} \rightarrow 0$ når $N \rightarrow \infty$ (Konsistens)
2. $\theta_0 - \hat{\theta} \rightarrow 0 \Rightarrow \psi_0 - \hat{\psi} \rightarrow 0$

Det første er et velkjent resultat fra asymptotisk teori og er direkte gitt fra egenskapene til mle under normale forhold. Et kritisk krav er om datasettet tilhører den parametriske familien f_{θ_x} eller ikke. Med andre ord, om den underliggende sannsynlighetsmodellen er korrekt. I situasjoner der dataene hører til den parametriske familien f_{θ_x} slik at

$$\sqrt{N}(\hat{\theta} - \theta_0) \sim N(0, \frac{1}{I(\theta_0)})$$

3.1. TEORETISK FEILANALYSE I

er asymptotisk normal og følgelig konsistent[5], hvor informasjonsmatrisen er gitt

$$I(\theta_0) = -E \left(\frac{\partial^2 \log(f(N_x/\theta_0))}{\partial \theta^2} \right)$$

Resultatet kan enkelt generaliseres til fler-parametrisk situasjon.

Derimot, i oppgaven er det allerede antatt at det parametriske rommet ligger utenfor korrekt dødelighet. Et naturlig spørsmål er hva som skjer med egenskapene til maximum likelihood estimatoren $\hat{\theta}$ under slike forhold. Peter J. Huber[12] viste at $\hat{\theta}$ er konsistent også når modellen ligger utenfor den sanne fordelingen. Ved å introdusere

$$Y = \frac{1}{\sqrt{N}} \sum_{x=1}^e \frac{\partial \log(f_{\theta_0}(N_x))}{\partial \theta}, \quad Z = \frac{1}{N} \sum_{x=1}^e \frac{\partial^2 \log(f_{\theta_0}(N_x))}{\partial^2 \theta}$$

og

$$I_{k,j}(\theta_0) = \text{var}[Y] = E \left(\left(\frac{\partial \log(f(N_x|\theta_0))}{\partial \theta_k} \right) \left(\frac{\partial \log(f(N_x|\theta_0))}{\partial \theta_j} \right) \right)$$

$$\frac{\partial}{\partial \theta_k} \lambda(\theta_0) = E[Z] = E \left(\frac{\partial}{\partial \theta_k} \frac{\partial}{\partial \theta_j} \log(f(N_x|\theta_0)) \right) \quad \text{for } k, j = 1, \dots, w$$

hvor $\Lambda = \lambda'_{kj}(\theta_0)$ og $I_g = I_{gjk}(\theta_0)$. Når dataene til modellen ligger utenfor det korrekte parameterrommet er

$$\sqrt{N}(\hat{\theta} - \theta_0) \sim N(0, \frac{I(\theta_0)}{(\lambda'(\theta_0))^2})$$

asymptotisk normal og følgelig konsistent. Det kan utvides til en fler-parametrisk situasjon slik at $\sqrt{N}(\hat{\theta} - \theta_0) \sim N(0, \Lambda^{-1} I_g \Lambda^{-1})$ der $\Lambda^{-1} I_g \Lambda^{-1}$ er kovariansmatrisen til estimatet $\hat{\theta}$ med dimensjon w X w der w er antall parametre i θ . Det gjenstår å argumentere for punkt 2. Hvis ψ og q_θ er kontinuerlige i alle punkter følger det at hvis

$$\hat{\theta} - \theta_0 \rightarrow 0 \Rightarrow q_{\hat{\theta}x} - q_{\theta_0x} \rightarrow 0 \Rightarrow \hat{\psi} - \psi_0 \rightarrow 0$$

slik at (2) holder.

En interessant problemstilling er å se om teorien ovenfor stemmer. Siden maximum likelihood estimatet er konsistent skal estimatet fra asymptotisk matematikk være tilsvarende bootstrap estimatet fra forrige seksjon. Det vil kunne gi en god indikasjon på hvor godt asymptotisk teori stemmer.

For en gitt parametrisk funksjon $f_{\theta x}$ definert i kapittel 2 er det mulig å skrive ut

3.2. TEORETISK FEILANALYSE II

resultatene med teorien ovenfor i en gitt livsforsikrings situasjon. Det er nødvendig å finne uttrykk for $\frac{\partial \log(f_{\theta x})}{\partial \theta}$ og $\frac{\partial^2 \log(f_{\theta x})}{\partial \theta^2}$ hvor

$$\log(f_{\theta x}) = \sum_l \left(\binom{N_x}{y_x} + y_x \log(q_{\theta x}) + (N_x - y_x) \log(1 - q_{\theta x}) \right)$$

slik at

$$\begin{aligned} \frac{\partial \log(f_{\theta x})}{\partial \theta_j} &= \sum_x \left(\frac{y_x \frac{\partial q_{\theta x}}{\partial \theta_j}}{q_{\theta x}} + \frac{(N_x - y_x) \frac{\partial q_{\theta x}}{\partial \theta_j}}{1 - q_{\theta x}} \right) \\ &= \sum_x \left(\frac{y_x - N_x q_{\theta x}}{q_{\theta x}(1 - q_{\theta x})} \right) \frac{\partial q_{\theta x}}{\partial \theta_j} \\ \frac{\partial^2 \log(f_{\theta x})}{\partial \theta_k \partial \theta_j} &= \sum_l \left(\left(\frac{y_x - N_x q_{\theta x}}{q_{\theta x}(1 - q_{\theta x})} \right) \frac{\partial^2 q_{\theta x}}{\partial \theta_k \partial \theta_j} + \frac{\partial q_{\theta x}}{\partial \theta_k} \cdot \frac{\partial}{\partial \theta_j} \left(\frac{y_x - N_x q_{\theta x}}{q_{\theta x}(1 - q_{\theta x})} \right) \right) \end{aligned}$$

Det gjenstår å ta forventningen til uttrykkene ovenfor. Siden y_x er antatt binomisk fordelt følger det at $E(y_x) = N_x g_x$ og $E(y_x)^2 = \text{Var}(y_x) + (E(y_x))^2$ hvor $\text{Var}(y_x) = N_x g_x(1 - g_x)$. Mer om dette i kapittel 4.

3.2 Teoretisk feilanalyse II

Resultatet fra forrige seksjon forteller at maximum likelihood estimatet $\hat{\theta}$ er konsistent og at avviket $\hat{\theta} - \theta_0$ går mot 0 når $N \rightarrow \infty$ hvis θ_0 er det optimale estimatet til korrekt dødelighet \underline{g} . Denne egenskapen kombiner med teorien som følger skal brukes til å analysere feil som skyldes at modellen ikke gjenspeiler virkeligheten eksakt. Kullback-Leibler distansen spiller en sentral rolle.

3.2.1 Kullback-Leibler distansen

Kullback-Leibler avstanden er et ikke-symmetrisk avstandsmål mellom to gitte sannsynlighetsfordelinger, G og F. I denne oppgaven brukes det kun diskrete sannsynlighetsmodeller slik at den diskrete KL avstanden er gitt[15]

$$D_{KL}(F_{\theta}|G) = \sum_x g_x \log\left(\frac{g_x}{f_{\theta x}}\right) \quad (3.1)$$

hvor typisk problemstilling er å la G være den sanne fordelingen, mens F er en modell/tilnærming/teori som skal beskrive G der (3.1) beskriver avviket/feilen.

3.2.2 Feil i antatt modell

Uansett størrelse på historisk data eller eksponering i en pensjonsportefølje, vil det alltid være feil i modellene som benyttes. Hvor stor feilen er og hvor godt modellen faktisk beskriver virkeligheten er essensielt. I forrige seksjon ble total feil representert ved å introdusere $\underline{q}_{\theta_0} = (q_{\theta_0 1}, q_{\theta_0 2}, \dots, q_{\theta_0 x_e})$ og $\psi_0 = \psi(\underline{q}_{\theta_0})$ slik at gjenstående feil dekomponeres

$$|\psi - \psi_{\hat{\theta}}| = |\psi - \psi_0 + \psi_0 - \psi_{\hat{\theta}}|$$

hvor det ble hevdet θ_0 er parameteren som determinerer dødeligheten \underline{q}_{θ_0} som ligger nærmest den korrekte dødeligheten \underline{g} . Dette er riktig siden θ_0 er parameteren som minimerer Kullback-Leibler distansen mellom to gitte sannsynlighetsfordelinger. I følge teorien gir dette den minste avstanden mellom korrekt og parameterisert dødelighet, og er essensielt i ideen som følger. Det er allerede vedtatt at maximum likelihood estimatet er konsistent slik at avviket $\hat{\theta} - \theta_0$ går mot 0 ved tilstrekkelig eksponering når θ_0 er det optimale estimatet. Det er ensbetydende med at avviket $\psi - \psi_0$ går mot 0. Siden feil uttrykkes i absolutte størrelser er

$$\begin{aligned} |\psi - \psi_{\hat{\theta}}| &= |\psi - \psi_0 + \psi_0 - \psi_{\hat{\theta}}| \\ &\leq \underbrace{|\psi - \psi_0|}_{\text{Systematisk feil}} + \underbrace{|\psi_0 - \psi_{\hat{\theta}}|}_{\text{Tilfeldig feil}} \end{aligned}$$

der $|\psi - \psi_0|$ er definert systematisk feil, mens $|\psi_0 - \psi_{\hat{\theta}}|$ er tilfeldig feil. Denne teorien kombinert med forrige seksjon leder oss til kjernen av problemstillingen. Anta at \underline{g} er **kjent**. Videre defineres en passende parametrisk fordeling f_{θ_x} for å determinere parametersettet θ_0 . Deretter simuleres bootstrap-estimatene slik at feilene kan evalueres separat. Dette gir mulighet til å analysere hvor mye eksponeringsstørrelse som er nødvendig for en pensjonsportefølje. Dessuten gir det grunnlag for å analysere modellfeil, altså feil som skyldes at modellen ikke gjenspeiler virkeligheten eksakt. En slik analyse vil indikere hvor godt dødelighetsmodellen fungerer som standardmodell i en gitt livsforsikringssituasjon.

En metode for å rapportere systematisk feil er å måle i forhold til *roten av kvadratisk feil*(RKF)³. Dette finnes ved å kvadrere uttrykket ovenfor og determinere forventningsverdien

$$\begin{aligned} E(\psi - \psi_{\hat{\theta}})^2 &= E(\psi - \psi_0 + \psi_0 - \psi_{\hat{\theta}})^2 \\ &= (\psi - \psi_0)^2 + E(\psi_0 - \psi_{\hat{\theta}})^2 - 2 \cdot (\psi - \psi_0) E(\psi_0 - \psi_{\hat{\theta}}) \end{aligned}$$

hvor $(\psi - \psi_0) E(\psi_0 - \psi_{\hat{\theta}})$ er restleddet. Da er [13]

$$\text{RKF} = \sqrt{E(\psi - \psi_{\hat{\theta}})^2}$$

³root-mean-square error(RMSE) på engelsk

3.2. TEORETISK FEILANALYSE II

slik at systematisk feil i prosent er definert

$$\text{Systematisk feil i prosent} = \frac{|\psi - \psi_0|}{\sqrt{E(\psi - \psi_{\hat{\theta}})^2}} \quad (3.2)$$

hvor $|\psi - \psi_0| = \sqrt{(\psi - \psi_0)^2}$. Restleddet inneholder systematisk feil slik at 3.2 ikke er helt teoretisk korrekt. Allikevel er det valgt å bruke denne representasjonen fordi restleddet vil gå mot 0 når N øker. Dette stemmer siden $E(\psi_0 - \psi_{\hat{\theta}})$ går mot 0 ved tilstrekkelig stor N. Det betyr at restleddet, relativt sett, har liten innflytelse når eksponeringsstørrelsen er stor.

Metoder for å studere feil i antatt modell

Mål	Periode	
	2009	1960 - 2008
Systematisk feil i prosent	$\frac{ \psi - \psi_0 }{\sqrt{E(\psi - \psi_{\hat{\theta}})^2}}$	$E_x \left(\frac{ \psi - \psi_0 }{\sqrt{E(\psi - \psi_{\hat{\theta}})^2}} \right)$
Relativ systematisk feil(RSF) i prosent	$\frac{ \psi - \psi_0 }{\psi}$	$E_x \left(\frac{ \psi - \psi_0 }{\psi} \right)$
Relativ tilfeldig feil(RTF) i prosent	$\frac{\sqrt{E(\psi_0 - \hat{\psi})^2}}{\psi}$	$E_x \left(\frac{\sqrt{E(\psi_0 - \hat{\psi})^2}}{\psi} \right)$
Relativ rot kvadratisk feil(RRKF) i prosent	$\frac{\sqrt{E(\psi - \hat{\psi})^2}}{\psi}$	$E_x \left(\frac{\sqrt{E(\psi - \hat{\psi})^2}}{\psi} \right)$

Tabell 3.1: Viser en oppsummering på diverse mål som brukes i forbindelse med modellfeilanalysen i kapittel 5

Systematisk feil i % vil være avgjørende i vurdering av modellfeil. I Kapittel 5 anvendes denne teorien for en rekke land. Siden dødelighetene innad i landene er vidt forskjellige er det ønskelig å vurdere datasettene opp mot hverandre og analysere avvik i den underliggende modellen. Det blir viktig å sammenligne feilene på riktig grunnlag, og følgende mål skal brukes

$$\begin{aligned} \text{Relativ systematisk feil} &= \frac{|\psi - \psi_0|}{\psi} \\ \text{Relativ tilfeldig feil} &= \frac{\sqrt{E(\psi_0 - \hat{\psi})^2}}{\psi} \\ \text{Relativ RKF} &= \frac{\sqrt{E(\psi - \hat{\psi})^2}}{\psi} \end{aligned}$$

I fler-periode analysen er det lite informativt å illustrere målene ovenfor med hensyn på alder og periode. Istedenfor vil resultatene bli presentert ved å ta forventningen med hensyn på alderen, og blir omtalt som *gjennomsnittelig relativ feil*. Tabell 3.1 viser en total oversikt over målene som blir brukt i kapittel 5.

Engangspremiefeil i antatt modell

En engangspremie i livsforsikring kjennetegner et beløp som blir betalt kun en gang i en bestemt periode⁴, og vil sikre et forhånsbestemt beløp som utbetales årlig livet ut fra pensjonsalder.

Ved gitt dødelighetsintensitet v er sannsynligheten for å overleve fra x til $x+k$ perioder gitt ved[1]

$${}_k p_x = \exp \left(\int_x^{x+k} -v dy \right)$$

og diskonteringsrenten er gitt ved $\frac{1}{1+r}$ hvor r er renta slik at en engangspremie er matematisk definert følgende[1]

$$\Psi = s \sum_{k=\max(x_r-x_0)}^{x_e-k} d_k^k p_x$$

Det gir en systematisk feil i engangspremier ved $\Psi - \Psi_0$, og målt i forhold til tilfeldig feil er

$$\text{Systematisk feil i prosent} = \frac{|\Psi - \Psi_0|}{\sqrt{E(\Psi - \Psi_{\hat{\theta}})^2}}$$

3.3 Pensjon- og dødelighetsfeil i lengre levetid

Det er komplisert⁵ å gjennomføre feilanalyse fra seksjon 3.1 og 3.2 med hensyn på Lee-Carter modellen. Bakgrunnen for det bygger på den stokastiske tidsrekken k_k som er en ikke lineær tilstandsmodell. Leseren henvises til [8] for ytterligere bakgrunn. Det utføres istedenfor to forskjellige feilevalueringer med hensyn på Lee-Carter modellen. Det første er å analysere *sensitiviteten* og *stabiliteten* til modellen. Dette gjøres ved å bruke forskjellige historiske dødelighetsperioder, og sammenligne økonomisk effekt gjennom engangspremier. Det andre er å utføre en *backtesting* av kjente dødeligheter. Ideen er hentet fra Dowd[7] og oppskriften er følgende

1. Definer *basisperioden* som skal brukes til å finne parametrene a_x , b_x og k_k , og deretter predikere dødelighetstrenden K år frem i tid, som allerede er kjent. Mer om det i kapittel 4.
2. Bruker parametrene til å predikere $m_{x,k+K}$, og beregner fremtidig forventet levealder for nyfødte. Det er et praktisk mål ettersom det tas i bruk alle dødelighetene for hver periode. Steg 1 og 2 kan eventuelt repeteres flere ganger
3. Til slutt brukes den allerede kjente og korrekte dødeligheten $g_{x,k}$

⁴Kan både betales før og etter pensjonstid

⁵sannsynligvis umulig

Kapittel 4

Datasett og Estimeringsmetoder

Historisk dødelighetsdata er nøkkelen til å estimere og beregne dødelighet i livsforsikring. Siden dødelighetene er i kontinuerlig forandring er det viktig med gode systemer som er lett tilgjengelig og hyppig oppdatert. Human Mortality database er et slikt system, og er den viktigste kilden til estimeringsmetodene som blir introdusert i seksjon 4.2, 4.3 og 4.4

4.1 Datasett

Human Mortality Database

Human Mortality Database[23] ble laget for å gi et kraftfullt samtidig en oversiktelig database på befolkning og dødelighetsdata til forskere, studenter, journalister, politikere og andre interessegrupper. Databasen ble opprettet i samarbeid mellom universitetet i Berkeley og institutt for demografisk forskning i Rostock, Tyskland. Prosjektet er finansiert over hele verden, og i tillegg gode avtaler med en rekke land som oppgir dataene. Antall land er begrenset siden det utelukkende aksepteres land med en komplett og detaljert oversikt fra en lengre periode. En slik omfattende oversikt koster penger, og det er idag registrert 37 land på Human Mortality database hvor ingen representerer Afrika.

Perioder og land

I kapittel 5 vil de numeriske resultatene bli presentert. Det er blitt brukt både en periode og flere perioder hendholdsvis året 2009 og årene 1960-2009.

R-kodene kan enkelt anvendes i begge scenarier med små endringer. I Lee-Carter modellen er det vesentlig å bruke historisk data i en lengre periode, og det blir brukt perioder fra 1930 helt opp til 2009. R-pakken *demography*[24] gjør det enkelt

4.1. DATASETT

Land	Periode	Alder
Alle ¹	1960 - 2008	0 - 105
Alle	2009	0 - 110

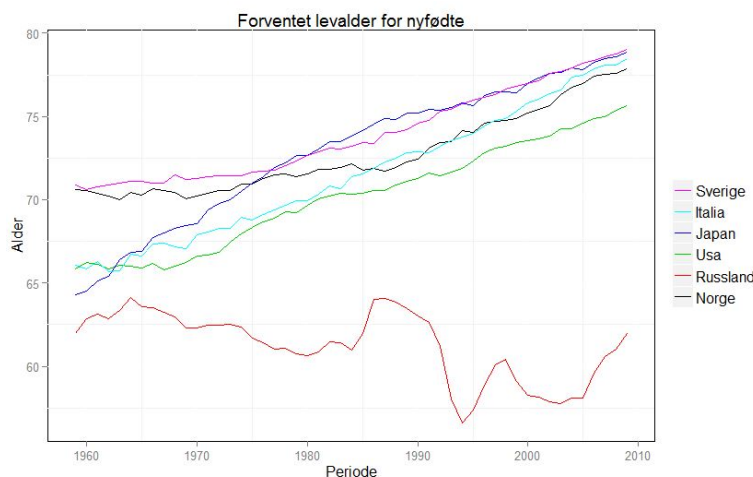
Tabell 4.1: Viser Datagrunnlag fra HMD

å laste ned dødelighetsdata fra Human mortality database, og er hyppig brukt i estimeringsmetodene som følger.

I utgangspunktet gjøres det analyser med hensyn på norske og russiske menn i periodene ovenfor. For ytterligere dokumentasjon, gjøres det beregninger for landene Italia, Japan, USA og Sverige. USA har et enormt multikulturelt samfunn og stort problem innenfor fedme og kosthold som påvirker dødeligheten. Det er stor forskjell innad i landet, med dødeligheter i pensjonskasser som er langt lavere enn landsgjennomsnittet. Italia har i overkant av 60 millioner innbyggere. Befolkningsveksten har skutt fart perioden 2000-2009. Italia har idag store økonomiske problemer, og det blir interessant å følge videre trend i befolkningen.

Japan er et av landene i verden med lavest dødelighet. Det ble i 2006 estimert at japanere lever gjennomsnittlig 81.25 år. Allikevel har det vært en tilbakegang i befolkningen ettersom det fødes relativt få barn. Idag ligger befolkningen på 128 millioner mennesker, og det er anslått at i 2060 vil være en nedgang til 87 millioner mennesker hvorav over 40 % er over 65 fylte år. Det betyr at det antakeligvis er flere pensjonister enn yrkesaktive om snaut 50 år i Japan. Det er en skremmende utvikling.

Kombinasjonen av forskjellige dødelighetsdata vil gi et godt innblikk hvordan modellen beskriver dødelighet generelt. Sverige er et nordisk land som har store likhetstrekk med Norge slik at det forventes tilsvarende resultater deretter.



Figur 4.1: Figur viser forventet levealder for nyfødte i perioden 1960-2009 for en rekke land.

Figur 4.1 oppsummerer de reelle utviklene til de nevnte landene ovenfor. Russland skiller seg tydelig ut, mens japanske menn er forventet å leve lengst.

4.2 Gompertz-Makeham

Fordelen (Sammenlignet med en ikke parametrisk modell) med Gompertz-Makeham modellen er at den kun beskrives av 3 parametere. Bruker antakelsen at hvert dødsfall er binomisk fordelt (n_x, g_x) hvor $n_l = N \cdot w_x$ er aldersprofilen definert i kapittel 2 slik at teorien ovenfor gir følgende parametrisk funksjon i alder x [1]

$$\begin{aligned} f_{\theta x} &= \prod_{x=1}^{x_e} \binom{n_x}{y_x} q_{\theta x}^{y_x} (1 - q_{\theta x})^{n_x - y_x} \\ &= \prod_{x=1}^{x_e} \binom{n_x}{y_x} (1 - \exp(-a - b \cdot e^{c \cdot x}))^{y_x} (\exp(-a - b \cdot e^{c \cdot x}))^{n_x - y_x} \end{aligned} \quad (4.1)$$

Maximum likelihood estimatorene $\hat{\theta} = (\hat{a}, \hat{b}, \hat{c})$ determineres ved å løse likningssystemet

$$\frac{\partial \log(f_{\theta x})}{\partial \theta} = 0$$

hvor

$$\log(f_{\theta x}) = \sum_x \log \binom{n_x}{y_x} + y_x \log(1 - \exp(-a - b \cdot e^{c \cdot x})) + (n_x - y_x) \log(\exp(-a - b \cdot e^{c \cdot x}))$$

Kullback-Leibler minimering

Den diskrete Kullback-Leibler distansen som skal minimeres med hensyn på θ_0 blir følgende

$$\begin{aligned} D_{KL}(F_{\theta}|G) &= \sum g_x \log\left(\frac{g_x}{f_{\theta_0 x}}\right) \\ &= \sum g_x (\log(g_x) - \log(f_{\theta_0 x})) \end{aligned}$$

hvor g_x er kjent. Dette er det samme som å minimere $-\sum g_x \log(f_{\theta_0 x})$ med hensyn på θ_0 . Kombinerer dette med uttrykket i (4.1) som gir

$$= - \sum_x g_x \log\left(\prod \binom{n_x}{y_x} q_{\theta x}^{y_x} (1 - q_{\theta x})^{n_x - y_x}\right)$$

Hvor $q_{\theta x}$ er Gompertz-Makeham dødeligheten. Bruker nå aldersprofilen definert ovenfor og det faktum at $E(y_x) = n_x \cdot g_x$ slik at

$$\begin{aligned} &\sum_x \sum_{g_x} \left(\binom{n_x}{y_x} + y_x \log(q_{\theta x}) + (1 - g_x) \log(1 - q_{\theta x}) \right) \\ &= -K - N \left(\sum_x w_x (g_x \log(q_{\theta x}) + (1 - g_x) \log(1 - q_{\theta x})) \right) \end{aligned}$$

4.2. GOMPERTZ-MAKEHAM

hvor N og K betydningsløse konstanter med tanke på minimeringen. Det viser at θ_0 er upåvirket av eksponeringsstørrelse N . Det gjenstår å maksimere følgende

$$\sum_x w_x (g_x \log(q_{\theta x}) + (1 - g_x) \log(1 - q_{\theta x}))$$

som enkelt implementeres numerisk.

Algoritme 4.2.1 og 4.2.2 viser pseudo kode henholdsvis for utregning av parametersettet θ_0 og engangspremie Ψ . Kodene vil bli brukt i forskjellige sammenhenger både i en eller flere perioder, og er grunnsteinen i mange numeriske utregninger i kapittel 5.

Algorithm 4.2.1: GOMPERTZ-MAKEHAM()

```

INPUT :  $g_x, w_x, \theta_{\text{start}}$ 
Gompertz-Makeham  $\leftarrow$  function( $\theta_{\text{start}}, g_x, w_x$ ){
   $t \leftarrow \exp(\theta_{\text{start}})$ 
   $l \leftarrow 1 : \text{length}(g_x)$ 
   $q_{\theta} \leftarrow 1 - \exp(-t[1] - t[2] \cdot \exp(t[3] \cdot l))$ 
   $-\sum_x w_x (g_x \log(q_{\theta x}) + (1 - g_x) \log(1 - q_{\theta x}))$ 
}
 $\theta_0 \leftarrow \text{optim}(\text{Gompertz-Makeham}(\theta_{\text{start}}, g_x, w_x))$ 
return ( $\theta_0$ )

```

Algorithm 4.2.2: ENGANGSPREMIE()

```

INPUT :  $q_{\theta_0 x}, x_e, x_r, s, d$ 
Engangspremie  $\leftarrow$  function( $q_{\theta}, x_r, s, d$ ){
   ${}_k p_x \leftarrow \text{cumprod}(1 - q_{\theta_0 x})$ 
   $\Psi \leftarrow s \sum_{k=\max(x_r, l_0)}^{x_e - k} d_k^k p_x$ 
}
return ( $\Psi$ )

```


Bootstrap simuleringer

Bootstrap ble introdusert i kapittel 2, og er verktøyet som blir brukt i forbindelse med å finne det tilfeldige avviket. Algoritme 4.2.3 viser en enkel pseudo kode for beregning av m_b bootstrap.

Algorithm 4.2.3: BOOTSTRAP()

```
INPUT :  $g_x, n_x \leftarrow w_x \cdot N, m_b$   
 $Z \leftarrow \text{matrix}(\text{rbinom}(m_b \cdot x, N_x, g_x), x, m_b))$   
for  $i \leftarrow 1$  to  $m_b$   
  do  $\begin{cases} \hat{\theta} \leftarrow \text{Gompertz-Makeham}(n_x, Z[, i]) \\ q_{\hat{\theta}} \leftarrow 1 - \exp(-\hat{\theta}[1] - \hat{\theta}[2] \cdot \exp(\hat{\theta}[3]x)) \\ \hat{\psi} \leftarrow \text{Engangspremie}(q_{\hat{\theta}}) \end{cases}$   
return ( $\hat{\psi}$ )
```

Koden utvides enkelt til en flerårig periode ved å kjøre algoritmen gjennom alle årene. Det eneste kravet er å implementere riktig historisk dødelighetsdata.

Akkumulert porteføljeverdi

Algoritme 4.2.4 viser pseudo-kode til en akkumulert pensjonsportefølje. Koden brukes i forbindelse med økonomiske eksperimenter i kapittel 5 og 6.

Algorithm 4.2.4: AKKUMULERT PORTEFØLJEVERDI()

```
INPUT :  $g_x, s, r, x_r, x_1, x_e, \Psi, N$   
 $\text{kpx} \leftarrow \text{cumprod}(1 - g_x[x_1 : x_2])$   
 $\text{gjenlevende} \leftarrow N \cdot \text{kpx}$   
 $\text{gjen.pens.alder} \leftarrow \text{gjenlevende}[-(1 : (x_r - x_1))]$   
 $\text{Vstart} \leftarrow \Psi[x_1] \cdot N$   
 $\text{Vinvestering} \leftarrow \text{Vstart} * (1 + r) ** (1 : (x_r - x_1))$   
 $\text{V.akk} \leftarrow \text{rep}(\text{Vstart}, \text{gjen.pens.alder}, \text{length}(x_r : x_e))$   
for  $i \leftarrow 1$  to  $(x_r - x_e)$   
  do  $\begin{cases} \text{if } (\text{V.akk}[i] > 0) \\ \text{V.akk}[i+1] \leftarrow (\text{V.akk}[i] - \text{gjen.pens.alder}[i] \cdot s) \cdot (1 + r) \\ \text{else} \\ \text{V.akk}[i+1] \leftarrow \text{V.akk}[i] - \text{gjen.pens.alder}[i] \cdot s \end{cases}$   
return ( $\text{V.akk}$ )
```

Asymptotiske utregninger

Det er nødvendig å finne $\frac{\partial \log(f_\theta)}{\partial \theta}$ hvor

$$\log(f_\theta) = \sum_x \left(\binom{y_x}{n_x} + y_x \log(q_{\theta x}) + (n_x - y_x) \log(1 - q_{\theta x}) \right)$$

som gir

$$\begin{aligned} \frac{\partial \log(f_\theta)}{\partial \theta} &= \sum_l \left(\frac{y_x \frac{\partial q_{\theta x}}{\partial \theta}}{q_{\theta x}} + \frac{(n_x - y_x) \frac{\partial q_{\theta x}}{\partial \theta}}{1 - q_{\theta x}} \right) \\ &= \sum_x \left(\frac{y_x - n_x q_{\theta x}}{q_{\theta x}(1 - q_{\theta x})} \right) \frac{\partial q_{\theta x}}{\partial \theta} \end{aligned}$$

hvor

$$\frac{\partial q_{\theta x}}{\partial \alpha} = 1 - q_{\theta x} \quad , \quad \frac{\partial q_{\theta x}}{\partial \beta} = \exp(cx)(1 - q_{\theta x}) \quad \text{og} \quad \frac{\partial q_{\theta x}}{\partial c} = \beta x \exp(cx)(1 - q_{\theta x})$$

som gir

$$\begin{aligned} \frac{\partial \log(f_{\theta x})}{\partial \alpha} &= \sum_x \frac{y_x - n_x q_{\theta x}}{q_{\theta x}} \quad , \quad \frac{\partial \log(f_{\theta x})}{\partial \beta} = \sum_x \frac{(y_x - n_x q_{\theta x}) \exp(cx)}{q_{\theta x}} \quad \text{og} \\ \frac{\partial \log(f_{\theta x})}{\partial c} &= \sum_l \frac{(y_x - n_x q_{\theta x}) \exp(cx) \beta x}{q_{\theta x}} \end{aligned}$$

De dobbeltderiverte blir nå

$$\begin{aligned} \frac{\partial^2 \log(f_{\theta x})}{\partial \alpha^2} &= - \sum_x \frac{y_x(1 - q_{\theta x})}{q_{\theta x}^2} \quad , \quad \frac{\partial^2 \log(f_{\theta x})}{\partial \beta^2} = - \sum_x \frac{y_x(\exp(2cx)(1 - q_{\theta x}))}{q_{\theta x}^2} \\ \frac{\partial^2 \log(f_{\theta x})}{\partial c^2} &= \sum_x \beta x^2 \exp(cx) \left(\frac{y_x - n_x q_{\theta x}}{q_{\theta x}} - \frac{y_x(1 - q_{\theta x})}{q_{\theta x}^2} \beta \exp(cx) \right) \end{aligned}$$

og

$$\begin{aligned} \frac{\partial^2 \log(f_{\theta x})}{\partial \alpha \partial \beta} &= - \sum_x \frac{y_x \exp(cx)(1 - q_{\theta x})}{q_{\theta x}^2} \quad \frac{\partial^2 \log(f_{\theta x})}{\partial \alpha \partial c} = - \sum_x y_x \left(\frac{\beta x \exp(cx)(1 - q_{\theta x})}{q_{\theta x}^2} \right) \\ \frac{\partial^2 \log(f_{\theta x})}{\partial \beta \partial c} &= \sum_x \exp(cx) x \left(\frac{y_x - n_x q_{\theta x}}{q_{\theta x}} - \beta y_x \exp(cx) \frac{1 - q_{\theta x}}{q_{\theta x}^2} \right) \end{aligned}$$

Det gir

$$\begin{aligned} \lambda_{1,1} &= - \sum_x n_x g_x \left(\frac{1 - q_{\theta x}}{q_{\theta x}^2} \right) \quad \lambda_{2,2} = - \sum_x n_x g_x \left(\frac{\exp(2c \cdot x)(1 - q_{\theta x})}{q_{\theta x}^2} \right) \\ \lambda_{3,3} &= \sum_x n_x \beta x^2 \exp(cx) \left(\frac{g_x - q_{\theta x}}{q_{\theta x}} - \frac{g_x(1 - q_{\theta x})}{q_{\theta x}^2} \beta \exp(cx) \right) \end{aligned}$$

4.3. BETA-BINOMISK MED DØDELIGHET

Videre er $\lambda_{1,2} = \lambda_{2,1}$, $\lambda_{1,3} = \lambda_{3,1}$ og $\lambda_{2,3} = \lambda_{3,2}$ slik at

$$\begin{aligned}\lambda_{1,2} &= - \sum_x n_x g_x \left(\frac{\exp(cx)(1 - q_{\theta x})}{q_{\theta x}^2} \right) & \lambda_{1,3} &= - \sum_x n_x g_x \left(\frac{\beta x \exp(cx)(1 - q_{\theta x})}{q_{\theta x}^2} \right) \\ \lambda_{2,3} &= \sum_x n_x \exp(cx) x \left(\frac{g_x - q_{\theta x}}{q_{\theta x}} - \beta g_x \exp(cx) \frac{1 - q_{\theta x}}{q_{\theta x}^2} \right)\end{aligned}$$

Tilsvarende må matrisen $I_{g,j,k}$ regnes ut hvor

$$\begin{aligned}v_{1,1} &= E_g \left(\frac{y_x - n_x q_{\theta x}}{q_{\theta x}} \right)^2 \\ &= \frac{1}{q_{\theta x}^2} E_g (y_x - n_x g_x + n_x (g_x - q_{\theta x}))^2 \\ &= \frac{1}{q_{\theta x}^2} (E_g (y_x - n_x g_x)^2 + E_g (n_x (g_x - q_{\theta x}))^2) \\ &= \frac{1}{q_{\theta x}^2} (n_x g_x (1 - g_x) + n_x^2 (g_x - q_{\theta x})^2)\end{aligned}$$

som følger mønsteret

$$I_{g,1,1} = \sum_x v_{1,1} \quad , \quad I_{g,2,2} = \sum_x v_{1,1} \exp(2cx) \quad \text{og} \quad I_{g,3,3} = \sum_x v_{1,1} \exp(2cx) \beta^2 x^2$$

og $I_{g,1,2} = I_{g,2,1}$, $I_{g,1,3} = I_{g,3,1}$ og $I_{g,3,2} = I_{g,2,3}$ slik at

$$I_{g,1,2} = \sum_x v_{1,1} \exp(cx) \quad , \quad I_{g,1,3} = \sum_x v_{1,1} \exp(cx) \beta l \quad \text{og} \quad I_{g,2,3} = \sum_x v_{1,1} \exp(2cx) \beta x$$

Algorithm 4.2.5: ASYMPOTISK()

INPUT : $g_x, q_{\theta 0x}, N_x \leftarrow w_x \cdot N$
Kovarianse $\leftarrow \Lambda^{-1} \cdot I_g \cdot \Lambda^{-1}$
Sd $\leftarrow \text{sqrt}(\text{diag}(\text{Kovarianse}))$

4.3 Beta-binomisk med dødelighet

For et gitt datasett $(y_1, y_2, \dots, y_{le})$ og $(n_1, n_2, \dots, n_{le})$ blir likelihood funksjonen $\ell(y)$ med modellen ovenfor følgende

$$\begin{aligned}\ell(y) &= \Pr(Y = y_1) \cdots \Pr(Y = y_{le}) \\ &= \binom{n_1}{y_1} \frac{\beta(y_1 + \gamma \xi_1, n_1 - y_1 + \gamma + \gamma \xi_1)}{\beta(\gamma \xi_1, \gamma + \gamma \xi_1)} \cdots \binom{n_{le}}{y_{le}} \frac{\beta(y_{le} + \gamma \xi_{le}, n_{le} - y_{le} + \gamma + \gamma \xi_{le})}{\beta(\gamma \xi_{le}, \gamma + \gamma \xi_{le})}\end{aligned}$$

Da blir Log-Likelihood funksjonen

$$\begin{aligned}\text{Log}(\ell(y)) &= \log \left(\binom{n_1}{y_1} \frac{\beta(y_1 + \gamma\xi_1, n_1 - y_1 + \gamma + \gamma\xi_1)}{\beta(\gamma\xi_1, \gamma + \gamma\xi_1)} \right) + \dots \\ &\quad + \log \left(\binom{n_{le}}{y_{le}} \frac{\beta(y_{le} + \gamma\xi_{le}, n_{le} - y_{le} + \gamma + \gamma\xi_{le})}{\beta(\gamma\xi_{le}, \gamma + \gamma\xi_{le})} \right) \\ &= \sum_1^{le} \left(\log \binom{n_i}{y_i} + \log \frac{\beta(y_i + \gamma\xi_i, n_i - y_i + \gamma - \gamma\xi_i)}{\beta(\gamma\xi_i, \gamma - \gamma\xi_i)} \right)\end{aligned}$$

For å finne parametrene $(\gamma, \xi_x = (\theta_0, \theta_1, \theta_2))$ må vi finne de parametrene som maksimerer $\ell(y)$, altså finne de parametrene som tilfredstiller $\frac{\partial \log(\ell(y))}{\partial \theta} = 0$. Siden $\log \binom{n_i}{y_i}$ er uavhengig av parametrene kan vi ekskludere dette leddet slik at

$$\begin{aligned}&= \sum_1^{le} \left(\log \frac{\beta(y_i + \gamma\xi_i, n_i - y_i + \gamma - \gamma\xi_i)}{\beta(\gamma\xi_i, \gamma - \gamma\xi_i)} \right) \\ &= \sum_1^{le} \log \left(\frac{\frac{\Gamma(y_i + \gamma\xi_i)\Gamma(n_i - y_i + \gamma - \gamma\xi_i)}{\Gamma(n_i + \gamma)}}{\frac{\Gamma(\gamma\xi_i)\Gamma(\gamma - \gamma\xi_i)}{\Gamma(\gamma)}} \right) \\ &= \sum_1^{le} (\log \Gamma(y_i + \gamma\xi_i) + \log \Gamma(n_i - y_i + \gamma - \gamma\xi_i) + \log \Gamma(\gamma) - \log \Gamma(n_i + \gamma) \\ &\quad - \log \Gamma(\gamma\xi_i) - \log \Gamma(\gamma - \gamma\xi_i))\end{aligned}\tag{4.2}$$

gjengstår. Ved få justeringer, er det mulig å bruke pseudokoden i seksjon 4.1. Forskjellen er antall parametre.

4.4 Lee-Carter

Estimeringsmetode

Den originale Lee-Carter modellen er definert ved

$$m_{x,k} = \exp(a_x + b_x \cdot k_k)\tag{4.3}$$

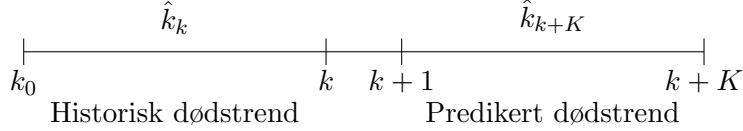
hvor (a_x, b_x) og k_k er henholdsvis alder og periodekomponenter. I den originale versjonen til Lee og Carter fra 1992 ble (4.3) estimeringsmetoden delt inn i følgende tre faser[17]

1. Tilpasse $\ln(m_{x,t}) = a_x + b_x k_k + e_{x,k}$ ved hjelp av singularverdi dekomposisjon(SVD) som løser minste kvadraters metode med hensyn på en definert

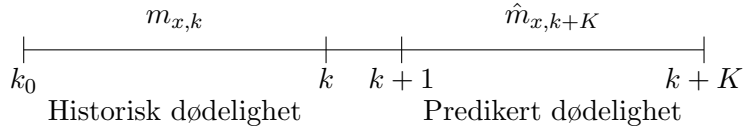
4.4. LEE-CARTER

basisperiode og bibetingelsene $\sum b_x = 1$ og $\sum k_k = 0$. Det gir $m_{x,k} \implies (\hat{a}_x, \hat{k}_k, \hat{b}_x)$. Justerer dødelighetstrenden ved å løse ligningen $D_{x,k} = N_{x,k} \cdot \exp(\hat{a}_x + \hat{b}_x \hat{k}_k)$ med hensyn på \hat{k}_k hvor $D_{x,k}$ er historisk antall døende i alder x og periode k.

2. Predikere dødelighetstrenden k_{k+K} ved hjelp av stokastiske tidsrekker



3. Predikere logaritme dødeligheter med $\hat{a}_x, \hat{k}_{k+K}, \hat{b}_x \implies \hat{m}_{x,k+K}$



I denne oppgaven brukes den bi-lineære poisson modellen definert i kapittel 2. Forskjellen i selve estimeringsmetoden ligger i fase 1, mens fase 2 og 3 er tilsvarende den originale Lee-Carter modellen ovenfor. Parametrene a_x , b_x og k_k finnes ved å maksimere log-likelihood funksjonen[3]

$$\ell(a_x, b_x, k_k) = \sum_x \sum_k (D_{x,k}(a_x + b_x \cdot k_k) - N_{x,k} \exp(a_x + b_x \cdot k_k))$$

siden leddet $b_x \cdot k_k$ er bilineært er det ikke mulig å estimere gjennom vanlig metode. Istedenfor brukes det en Newton-Raphson algoritme. Goodmann(1979) var den første til å foreslå denne itererende metoden for å estimere bi-lineære modeller. Algoritmen er definert følgende i iterasjon v+1[3]

$$\hat{\theta}^{(v+1)} = \hat{\theta}^{(v)} - \frac{\frac{\partial \ell^{(v)}}{\partial \theta}}{\frac{\partial^2 \ell^{(v)}}{\partial \theta^2}}$$

hvor $\theta = (a_x, b_x, k_k)$. Iterasjonen stopper når $\ell(a_x^{i+1}, b_x^{i+1}, k_k^{i+1}) - \ell(a_x^i, b_x^i, k_k^i) \leq 10^{-6}$. De partiellderivate er gitt[3]

$$\begin{aligned} \frac{\partial \ell}{\partial a_x} &= \sum_k (D_{x,k} - N_{x,k} \exp(a_x + b_x \cdot k_k)), & \frac{\partial^2 \ell}{\partial a_x^2} &= - \sum_k N_{x,k} \exp(a_x + b_x \cdot k_k) \\ \frac{\partial \ell}{\partial b_x} &= \sum_k k_k (D_{x,k} - N_{x,k} \exp(a_x + b_x \cdot k_k)), & \frac{\partial^2 \ell}{\partial b_x^2} &= - \sum_k k_k^2 \cdot N_{x,k} \exp(a_x + b_x \cdot k_k) \\ \frac{\partial \ell}{\partial k_k} &= \sum_x b_x (D_{x,k} - N_{x,k} \exp(a_x + b_x \cdot k_k)), & \frac{\partial^2 \ell}{\partial k_k^2} &= - \sum_x b_x^2 \cdot N_{x,k} \exp(a_x + b_x \cdot k_k) \end{aligned}$$

4.4. LEE-CARTER

Det er nå enkelt å skrive ut iterasjonen med startverdier $a_x = 0$, $b_x = \frac{1}{K}$ og $k_k = 0$ slik at [3]

$$\begin{aligned} a_x &= a_x + \frac{\sum_k (D_{x,k} - N_{x,k} \exp(a_x + b_x \cdot k_k))}{\sum_k N_{x,k} \exp(a_x + b_x \cdot k_k)} \\ k_k^* &= \hat{k}_k + \frac{\sum_x b_x (D_{x,k} - N_{x,k} \exp(a_x + b_x \cdot k_k))}{\sum_x b_x^2 \cdot N_{x,k} \exp(a_x + b_x \cdot k_k)} \\ k_k &= k_k^* - \text{mean}(k_k^*) \\ b_x &= b_x + \frac{\sum_k k_k (D_{x,k} - N_{x,k} \exp(a_x + b_x \cdot k_k))}{\sum_k k_k^2 \cdot N_{x,k} \exp(a_x + b_x \cdot k_k)} \end{aligned}$$

og for hver iterasjonen oppdateres likelihoodfunksjonen ℓ gitt ovenfor. 4.4.1 viser pseudo koden i R for å estimere a_x , b_x og k_k . Iterasjonen løper så lenge verdien til likelihood funksjonene mellom periode i og $i+1$ er større enn 10^{-6} . Videre oppdateres bibetingelsene $\sum b_x = 1$ og $\sum k_k = 0$ for å oppfylle estimeringskravet fra den originale versjonen. Ilc[4] er en R-pakke som gjør det enkelt å regne ut steg 1.

Algorithm 4.4.1: LEE-CARTER()

```

INPUT :  $D_{x,t}, N_{x,t}, a_x, b_x, k_k, i \leftarrow 0$ 
 $L \leftarrow \text{function}(\hat{a}_x, \hat{b}_x, \hat{k}_k) \{$ 
 $\sum_x \sum_k (D_{x,k}(a_x + b_x \cdot k_k) - N_{x,k} \exp(a_x + b_x \cdot k_k))$ 
 $\}$ 
while  $L(i+1) - L(i) > 10^{-6}$ 
   $i \leftarrow i + 1$ 
  do  $\begin{cases} \hat{a}_x \leftarrow \hat{a}_x + \frac{\sum_k (D_{x,k} - N_{x,k} \exp(a_x + b_x \cdot k_k))}{\sum_k N_{x,k} \exp(a_x + b_x \cdot k_k)} \\ k_k^* \leftarrow \hat{k}_k + \frac{\sum_x b_x (D_{x,k} - N_{x,k} \exp(a_x + b_x \cdot k_k))}{\sum_x b_x^2 \cdot N_{x,k} \exp(a_x + b_x \cdot k_k)} \\ \hat{k}_k \leftarrow k_k^* - \text{mean}(k_k^*) \\ \hat{b}_x \leftarrow b_x + \frac{\sum_k k_k (D_{x,k} - N_{x,k} \exp(a_x + b_x \cdot k_k))}{\sum_k k_k^2 \cdot N_{x,k} \exp(a_x + b_x \cdot k_k)} \end{cases}$ 
 $\hat{b}_x \leftarrow \frac{\hat{b}_x}{\text{sum}(\hat{b}_x)}$ 
 $\hat{k}_k \leftarrow \hat{k}_k \cdot \hat{b}_x$ 
return  $(\hat{a}_x, \hat{b}_x, \hat{k}_k)$ 

```

Deretter brukes stokastiske tidrekker til å predikere dødelighetstrenden i K år frem i tid. La [16]

$$k_{k+K} = k_{k+K-1} + \mu + \epsilon$$

hvor $\mu = \frac{k_k - k_{k_0}}{k-1}$ og $\epsilon \sim N(0, \sigma^2)$ hvor $\sigma^2 = \frac{1}{k-1} \sum_{k-1} ((k_{k-1} - k_k) - \mu)^2$. Dette gjøres m_b ganger. Algoritme 4.4.2 viser den avsluttende fasen av estimeringen.

Algorithm 4.4.2: LEE-CARTER()

INPUT : $\hat{a}_x, \hat{b}_x, \hat{k}_k, \mu, \sigma, m_b$
 $\hat{k}_{k+K} \leftarrow \hat{k}_{k+K-1} + \mu + \text{rnorm}(m_b, 0, \sigma)$
 $\hat{k}_{k+K} \leftarrow \text{mean}(\hat{k}_{k+K})$
 $\hat{m}_{x,k+K} \leftarrow \exp(\hat{a}_x + \hat{b}_x \cdot \hat{k}_{k+K})$
 $\hat{e}_{0,x,k+K} \leftarrow 0.5 + \text{apply}(\text{apply}(1 - \hat{m}_{x,k+K}, 2, \text{cumsum}), 2, \text{sum})$
return $(\hat{m}_{x,k+K}, \hat{e}_{0,x,k+K})$

Dynamisk dødelighet og engangspremier

Det må gjøres enkelte justeringer for å regne ut engangspremier med dynamisk dødelighet. Algoritme 4.4.3 viser pseudo koden. Siden dødeligheten endrer seg for hver periode er det nødvendig å oppdatere dødeligheten deretter. Det gjøres ved å bruke en for-løkke for hver alder fra startfasen x_1 til maksalder x_e . Selve oppdateringen gjøres ved å plukke ut diagonalelementene i matrisen $m_{x:x_e,k}$. Diagonalelementene blir deretter i hver runde brukt til å engangspremien i alder x for hver runde i loopen. Sluttresultatet gir en engangspremie vektor med aldrene x_1 opp til x_e

Algorithm 4.4.3: DYNAMISK ENGANGSPREMIE()

INPUT : $m_{x,k}, x_e, x_r, s, d$
Engangspremie $\leftarrow \text{function}(m_{x,k}, x_r, s, d)\{$
for $x \leftarrow x_1$ **to** x_e
 do $\begin{cases} K \leftarrow x \text{ to } x_e \\ {}_k p_K \leftarrow \text{cumprod}(\text{diag}(1 - m_{K,k})) \\ \Psi(x) \leftarrow s \sum_{k=\max(x_r-x)}^{x_e-k} d_k^k p_K \end{cases}$
}
return (Ψ)

Del II

Numeriske resultater

Kapittel 5

Gompertz-Makeham

Estimeringsfeil er sensitiv til pensjonsporteføljer med lav eksponeringsstørrelse. Det fører til feil i parameterene som igjen fører til feil i pensjonspremien. Ved tilstrekkelig eksponering i en portefølje vil feilen gå mot 0 slik at modellfeilen gjenstår. Det ble vist i kapittel 2. Hvor stor må en pensjonsportefølje være for at modellfeilen dominerer, og hvor stor denne feilen er relativ til hva en engangspremie koster idag er sentrale problemstillinger i kapittel 5.

I dette kapittelet gjøres det analyser med det ovennevnte for Gompertz-Makeham modellen på bakgrunn av en eller flere perioder. En periode tillater dypere og mer detaljerte studier, mens fler-periodene gir et bedre innblikk i trendene og den kontinuerlige styrken til modellen. Det brukes dødelighetsdata henholdsvis til en og flere perioder fra 2009 og 1960-2008.

5.1 Avvik i parametre og modell

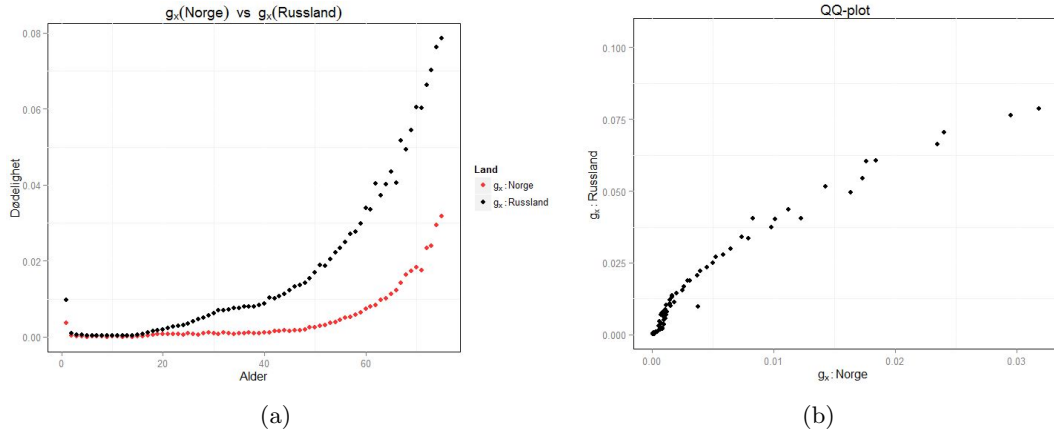
5.1.1 År 2009

	g_1	g_{10}	g_{30}	g_{50}	g_{75}
Russland	0.96 %	0.032 %	0.62 %	1.68 %	7.85 %
Norge	0.37 %	0.012 %	0.094 %	0.25 %	3.18 %

Tabell 5.1: Korrekt dødelighet i diverse aldre for norske og russiske menn.

Figur 5.1 a) viser korrekt dødelighet for norske og russiske menn fra 2009 i alderen 1 til 75 år. Ikke overraskende er det store avvik mellom landene. Veksten stemmer godt overens med den eksponentielle antakelsen i Gompertz-Makeham modellen. Forøvrig er veksten for russiske menn større. Det blir interessant å studere avvikene i modellen mellom landene med så store forskjeller vist ovenfor. Figur 5.1 b) viser et QQ-plot og det hvor opplagt ikke er 1-1 korrespondanse mellom landene. Forholdet mellom

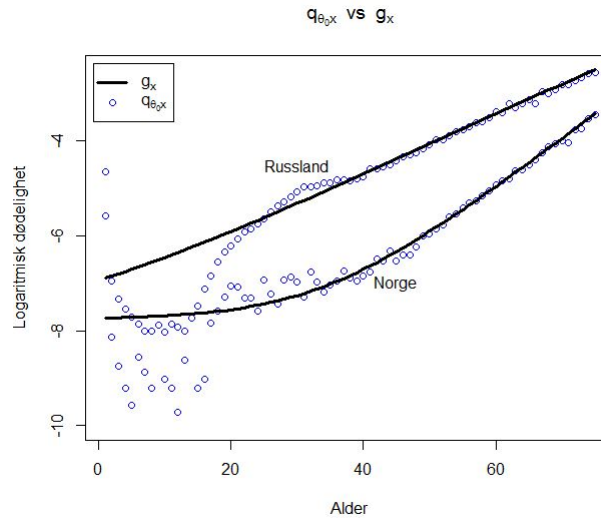
5.1. AVVIK I PARAMETRE OG MODELL



Figur 5.1: Figur a) viser korrekt dødelighet g_x for russiske og norske menn. Figur b) viser QQ-plot. Merk forskjellig akser i QQ-plot.

landene er størst i perioden 30 til 50 år, og g_{50} er nesten 7 ganger større for russiske menn. Se tabell 5.1.

Figur 5.2 illustrerer korrekt dødelighet plottet mot parametrisk Kullback-Leibler dødelighet q_{θ_0} for norske og russiske menn i alderen 1 til 75 år. Dødelighetene er transformert til logaritmisk skala, og ikke uventet er $\underline{g}(\text{Russland})$ mer lineær sammenlignet med $\underline{g}(\text{Norge})$. Årsaken til dette er i tråd med det såkalte fenomenet *rektalisering* som Pittaco[3] introduserte. Begrepet stammer fra den kraftige dødelighetsreduksjonen til de industrielle landene. Dødeligheten til norske menn i figur 5.1 viser en slik rektalisering, spesielt sammenlignet med russiske menn.



Figur 5.2: Korrekt dødelighet g_x plottet mot dødeligheten til det parametersettet θ_0 som ligger nærmest, for russiske og norske menn på logaritmisk skala

5.1. AVVIK I PARAMETRE OG MODELL

Fra alderen 20 til 75 år tilpasser q_{θ_0} meget bra for begge land. Det betyr at q_{θ_0} (Russland) er følgelig mer lineær. Tabell 5.2 viser utvalgte dødeligheter. Modellen fanger hverken opp den høye dødeligheten for nyfødte, og den lave dødeligheten som følger i alderen 2 til 20 år. Figuren ovenfor viser dette tydelig. Fra alderen 30 til 75 tilpasser modellen seg bedre, og forskjellen mellom g_{50} og $q_{\theta_0 50}$ er lav for begge land.

	$q_{\theta_0 1}$	$q_{\theta_0 10}$	$q_{\theta_0 30}$	$q_{\theta_0 50}$	$q_{\theta_0 75}$
Russland	0.102 %	0.157 %	0.493 %	1.71 %	8.32 %
Norge	0.044 %	0.046 %	0.071 %	0.27 %	3.23 %

Tabell 5.2: Diverse Kullback-Leibler dødelighet for norske og russiske menn.

Totalt sett viser Figur 5.2 indikasjon på glattere kurver ved bruk av russisk data, men vanskelig å gi håndfaste konklusjoner. Parameterene $\theta_0 = (a, b, c)$ er estimert til følgende verdier

$$\begin{aligned}\theta_0 &= (0.000315, 0.000657, 0.06507) \quad (\text{Russland}) \\ \theta_0 &= (0.000427, 0.000011, 0.10577) \quad (\text{Norge})\end{aligned}$$

hvor det er opplagt større forskjeller i alderskomponenten¹ med parameterene b og c (som beskriver den eksponentielle veksten i dødelighetene) enn komponenten som er uavhengig av alderen². Det er i tråd med det som ble diskutert ovenfor slik at en høyere c for norske menn er forventet ettersom dødelighetsveksten er lavere. Siden dødeligheten til norske menn er jevnt lavere for alle aldre er det nødvendig med en lavere b for å kompensere at dødeligheten forblir lavere i de yngre aldre. Det er mindre forskjell i Makeham-leddet hvor a (Norge) er størst. Det stemmer godt overens med tidligere funn, og er rimelig siden norske menn har lavere dødelighet uavhengig av alder

θ_0 er parametersettet til Gompertz-Makeham dødeligheten som ligger nærmest det korrekte dødelighetssettet. I den virkelige verden er ikke dette parametersettet tilgjengelig for et forsikringselskap, spesielt hvis datamengden eller eksponeringsstørrelsen er begrenset. Som nevnt i kapittel 3 skal det brukes bootstrap med maximum likelihood-estimator på en rekke eksponeringsstørrelser. I følge teorien vil $\hat{\theta} \rightarrow \theta_0$.

N = 1 000 000	E(\hat{q}_{θ_1})	E($\hat{q}_{\theta_{10}}$)	E($\hat{q}_{\theta_{30}}$)	E($\hat{q}_{\theta_{50}}$)	E($\hat{q}_{\theta_{75}}$)
Russland	0.112 %	0.166 %	0.498 %	1.714 %	8.32 %
Norge	0.045 %	0.047 %	0.071 %	0.273 %	3.23 %

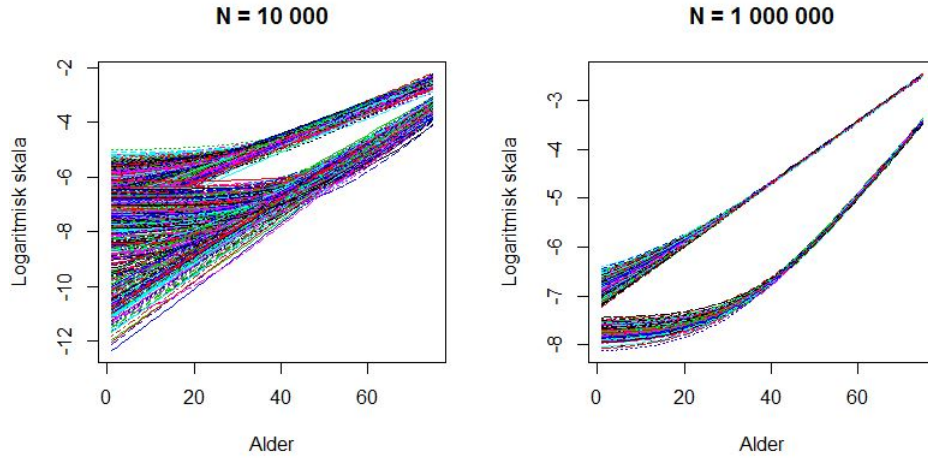
Tabell 5.3: Forventet dødelighet ved 500 bootstrap simulering ved N = 1 000 000, norske og russiske menn.

Figur 5.3 viser 500 bootstrap simuleringer med eksponeringsstørrelse på henholdsvis ti tusen og en million mennesker for norske og russiske menn. Det er stor forskjell

¹Gompertz-leddet

²Makeham-leddet

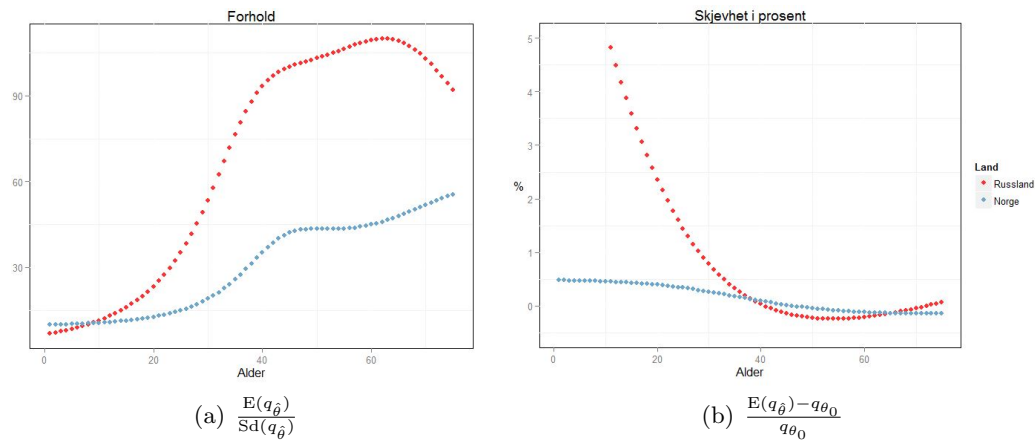
5.1. AVVIK I PARAMETRE OG MODELL



Figur 5.3: $m_b = 500$ bootstrap simuleringer ved $N = 10\,000$ og $N = 1\,000\,000$ for russiske og norske menn. Store forskjeller!

mellom størrelsene og risikoen er opplagt lavere når simuleringen løper på en million mennesker. Spredningen på simuleringene er større for norske menn. Tabell 5.3 viser den forventede dødeligheten ved 500 bootstrap simuleringer ved en million mennesker. Det er minimale avvik sammenlignet med q_{θ_0} i tabell 5.2 slik at konvergensen er tydelig.

Figur 5.4 a) viser forholdet mellom forventning og standardavvik til bootstrap simuleringene. Et høyere forhold impliserer mindre usikkerhet. Begge landene har et relativt lavt forhold i alderen 1-20 år, men vokser i ulikt tempo frem mot alder 50 år. Figuren viser at den relative usikkerheten er lavest i alderen 30 til 75 år, og det stemmer godt med tidligere funn (figur 5.2). Ikke like opplagt er hvorfor forholdet på russiske menn er høyere. Det er spesielt høyest i de aldrene det er størst prosentvis forskjell på bootstrap estimatorene mellom landene. Se tabell 5.3.



Figur 5.4: Venstre graf viser Forholdet mellom forventning og standardavvik til bootstrap simuleringene. Høyre graf viser den prosentmessige skjevheten mellom bootstrap simuleringen og Kullback-Leibler dødeligheten.

5.1. AVVIK I PARAMETRE OG MODELL

Figur 5.4 b) viser skjevheten mellom bootstrap og Kullback-Leibler dødeligheten. Det er en tydelig overestimering, som går mot 0 når alderen øker. Det er store forskjeller mellom landene, og spesielt i alderen 1 til 30 år med en skjevhet på hele 10 % på russiske nyfødte. I alder 40 til 60 år underestimerer russisk dødelighet. Norsk dødelighet konverger stabilt, og det er opplagt mindre skjevheter i aldrene 30 til 75 år. For å kunne si noe ytterligere er det nødvendig å analysere bootstrap-parameterene a, b og c.

Den forventede verdien til de tre parameterene ved en million mennesker er henholdsvis beregnet til å være

$$E(\hat{\theta}^*) = (0.0004319, 0.0006413, 0.0654) \quad (\text{Russland})$$

$$E(\hat{\theta}^*) = (0.0004331, 0.0000116, 0.1058) \quad (\text{Norge})$$

Sammenlignet med θ_0 er det minimale avvik utenom i parameteren a(Russland). Tabell 5.4 viser den prosentmessige skjevheten mellom $\hat{\theta}$ og θ_0 for parametrene a, b og c. Bootstrap-estimatet $E(\hat{\alpha})$ overestimerer med hele 36.8 %. Det er årsaken til den høye skjevheten for russiske menn i figur 5.4 venstre graf ettersom a er mer sensitiv i de lavere aldrene. Parameteren b er underestimert med -2.4 %, og er årsaken til den negative skjevheten mellom dødelighetene i figur 5.4. De øvrige skjevhetene er minimale, spesielt parameterene med hensyn på norske menn. Oppsummert er det større skjevheter ved russiske menn og er grunnen til mer ustabile resultater i figur 5.4.

Tabell 5.4 viser på venstre side forholdet mellom forventet verdi og standardavvik for bootstrap-estimatene. Ikke overraskende er forholdet lavest ved $\frac{E(\hat{a})}{Sd(\hat{a})}(\text{Russland}) = 2.4$, som er omtrent en fjerdedel sammenlignet mot tilsvarende parameter for norske menn. Videre er det lavere usikkerhet i parameter \hat{b} med hensyn på russiske menn. Parameteren \hat{c} har et høyt og nesten identisk forhold mellom landene, og har i tillegg den klart laveste forventningskjevheten mellom parameterene. Det betyr at c er den enkleste parameteren å estimere i Gompertz-Makeham modellen.

$m_b = 500$	N = 1 000 000				
Forhold	Russland	Norge	Skjevhet	Russland	Norge
$\frac{E(\hat{a})}{Sd(\hat{a})}$	2.3	9.6	$\frac{E(\hat{a})-a_0}{a_0}$	36.8 %	1.3 %
$\frac{E(\hat{b})}{Sd(\hat{b})}$	19.8	11.9	$\frac{E(\hat{b})-b_0}{b_0}$	-2.4 %	-0.2 %
$\frac{E(\hat{c})}{Sd(\hat{c})}$	92.6	103.1	$\frac{E(\hat{c})-c_0}{c_0}$	0.50 %	0.05 %

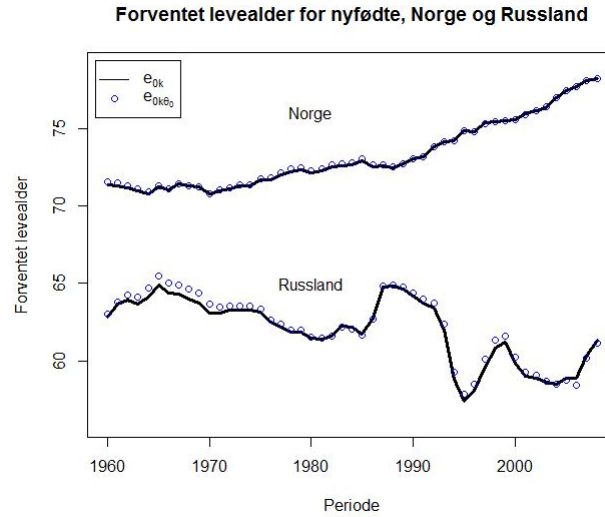
Tabell 5.4: Forholdet mellom forventningsverdi og standardavvik for parameterene α, β og c. Det er utført på grunnlag av 150 bootstrap simuleringer. Høy verdi indikerer liten usikkerhet. Venstre tabell viser skjevhet i % hvor negativ verdi følger av underestimering

Tabell 5.4 forklarer skjevheten som ble vist i bootstrap-dødelighetene ovenfor. Parameterene a og b forårsaker henholdsvis under og over-estimeringen for russiske menn i alderen 0 til 60 år. Videre er forholdet $\frac{E(\hat{b})}{Sd(\hat{b})}(\text{Russland})$ nesten dobbelt sammenlignet med Norge, og er trolig årsaken til at forholdet i dødeligheten er høyere,

5.1. AVVIK I PARAMETRE OG MODELL

spesielt i alderen 20 til 75 år. Det viser igjen at innflytelsen på Gompertz-Makeham parameterene er svært aldersavhengige.

5.1.2 Periode 1960 til 2008



Figur 5.5: Korrekt forventet levealder for nyfødte i perioden 1960-2008 plottet mot parametersettet som ligger nærmest. Den systematiske feilen er minimal hvor differansen er størst for russiske menn på 60-tallet og årene rundt 2000.

Figur 5.5 illustrerer henholdsvis korrekt og parametrisert(KL) forventet levealder på nyfødte i perioden 1960-2008. Avvikene er minimale og det er opplagt at Gompertz-Makeham beskriver perioden godt.

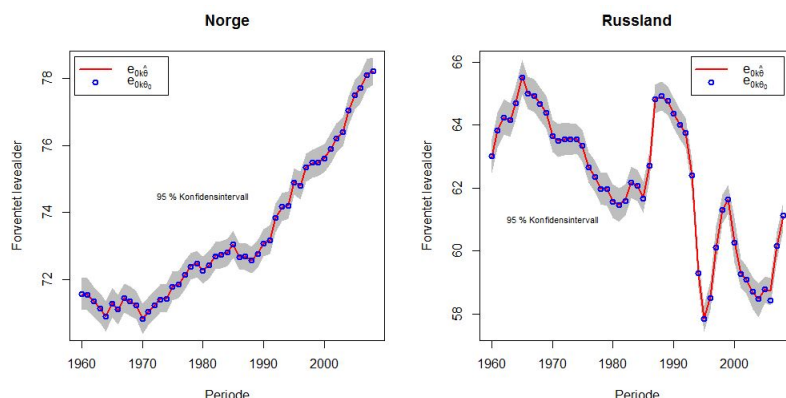
k	1970		1990		2009	
	Norge	Russland	Norge	Russland	Norge	Russland
$e_{0,k}$	71.22	63.73	72.76	64.62	78.33	61.78
$e_{\theta_0,0,k}$	71.22	64.39	72.70	64.75	78.28	61.12

Tabell 5.5

Allikevel illustrerer figuren perioder hvor modellen beskriver norsk fremtidig levealder bedre. Spesielt i perioden 1960 til 1970 og årene rundt 2000. Tabell 5.5 viser et utdrag fra figuren, og forventet levealder til norske menn er ekstremt presist. Mellom landene, er det langt større ulikheter i periodene 1970 og 2009³ enn 1990. Det kommer tydelig frem i figuren. Dessuten viser året 2009 et av de største avvikene mellom $e_{0,k}$ og $e_{0,2009,k}$. Det forsvaret Gompertz-Makeham modellen anvendt på russisk data, og vitner til mer presise resultater enn det som ble antatt i ett-år analysen fra 2009 ovenfor.

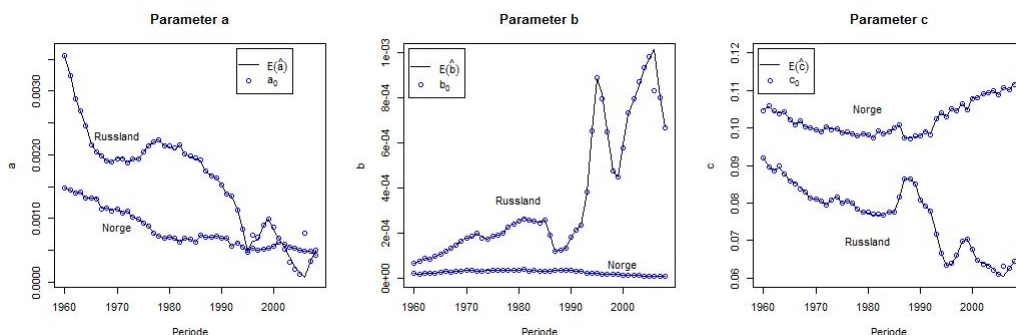
³Beregnet fra ett-år perioden

5.1. AVVIK I PARAMETRE OG MODELL



Figur 5.6: Viser 95 % konfidensintervall sammen med $e_{\theta,0,k}$ og $e_{\hat{\theta},0,k}$. Kovergensen kommer tydelig frem. Det er utført 500 bootstrap-simuleringer

Figur 5.6 viser forventet levealder ved 500 bootstrap kjøringar med eksponeringsstørrelse på en million mennesker og et tilhørende 95 % konfidensintervall. Grafene viser stor nøyaktighet hvor $e_{0k\hat{\theta}}$ konvergerer mot $e_{0k\theta_0}$ slik at den tilfeldige feilen ved forventet levetid er minimal. Dette er ikke overraskende og stemmer godt overens med resultatene fra året 2009. Mer om dette i pensjonsberegninger neste seksjon.



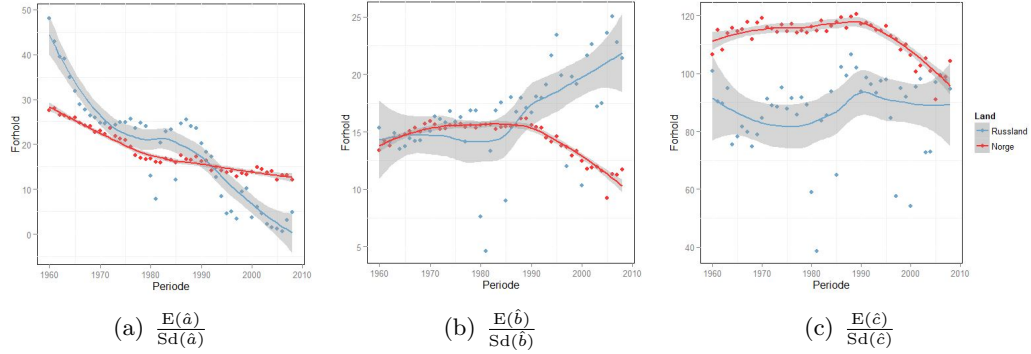
Figur 5.7: Viser Gompertz-Makeham parameterene $\hat{\theta}$ og θ_0 for norske og russiske menn i perioden 1960 til 2009. Forskjellene er store, og parameterene med hensyn på norske menn er mer stabilt!

Resultatene ovenfor viser større svingninger for russiske menn. Årsaken ligger naturlig nok i landets dødelighetsutvikling som påvirker parameterestimeringen i mye større grad. Figur 5.7 viser utviklingen til Gompertz-Makeham parameterene a, b og c i perioden 1960-2008. Det er ingen tvil at a, b og c(Russland) er mer ustabile. Parameter a har en nedadgående trend for begge land. a hører til Gompertz-Makeham leddet som er uavhengig av alderen. Det betyr at modellen vektlegger leddet i mindre grad. Gavrilov(2009) har tidligere vist[9] at denne komponenten reduseres når dødelighetstrenden avtar. Det stemmer godt overens for norsk data, men ikke for russisk.

Det interessante med studiet var at parameterene b og c holdt seg forholdsvis rolig. Figur 5.7 viser at det gjelder for norske menn og ikke russiske. Det er stor variasjon

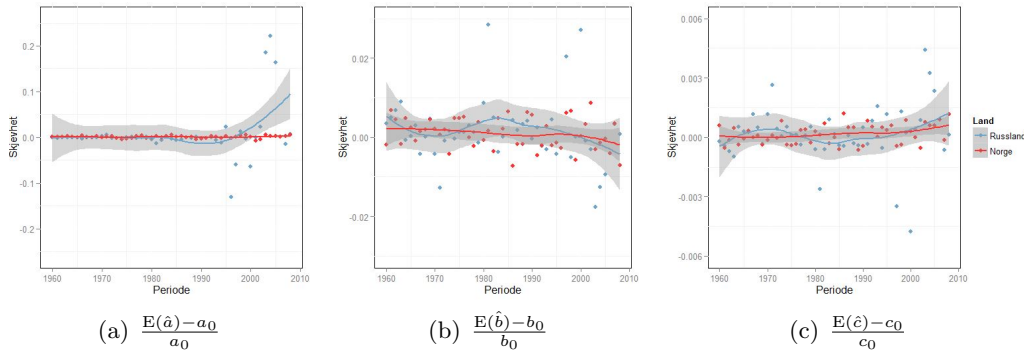
5.1. AVVIK I PARAMETRE OG MODELL

med uforutsigbar dødelighetstrend, spesielt etter 1990. Legg merke til at grafen til parameter c følger kurven til forventet levealder i figur 5.6. Det viser hvor stor påvirkning parameteren har i modellen. En høyere parameterverdi resulterer i en lavere dødelighet. Videre er det større avvik totalt sett mellom θ_0 og $E(\hat{\theta})$ for russiske menn.



Figur 5.8: Viser forholdet mellom forventning og standardavvik til parameterene a , b og c . $N = 1\,000\,000$

Figur 5.8 viser forholdet mellom forventning og standardavvik til parameterene \hat{a} , \hat{b} og \hat{c} med en eksponeringsstørrelse på en million mennesker. Det er tilpasset en ikke-parametrisk⁴ regresjon til punktene. Arealet rundt regresjonen viser et 95 % konfidensintervall. Totalt sett varierer forholdet mer for russiske menn, og konfidensintervallet er i alle tre situasjonene større. Forøvrig er forskjellen mellom landene større mot slutten av perioden, som allerede er analysert ovenfor (året 2009). Usikkerheten er lavest i parameteren c slik at denne er enklest å estimere. Det er tydelig at utviklingen i forholdet $\frac{E(\hat{\theta})}{sd(\hat{\theta})}$ (Norge) avtar ettersom årene går. Det betyr at usikkerheten øker i bootstrap simuleringene. Spørsmålet er hvor stor innflytelse det har på en pensjonsportefølje. Mer om det i neste seksjon.



Figur 5.9: Viser skjevhet mellom KL og bootstrap til parameterene a , b og c . $N = 1\,000\,000$

⁴Local Regression, tilpasser ved hjelp av R-pakken *ggplot2* [22]

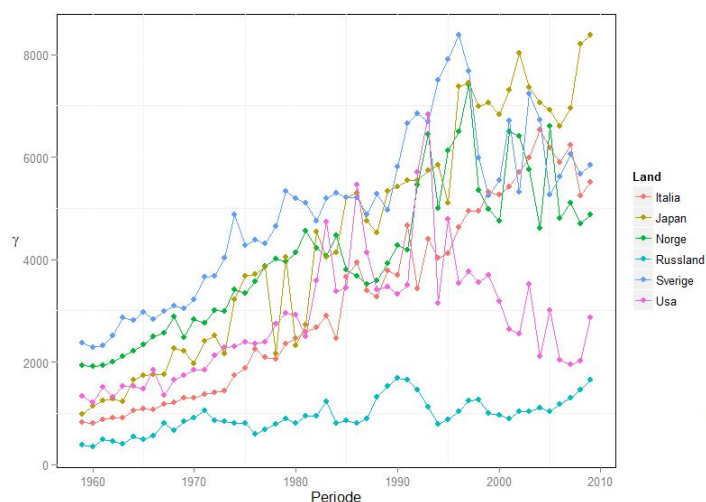
5.1. AVVIK I PARAMETRE OG MODELL

Figur 5.9 viser skjevheten mellom θ_0 og $\hat{\theta}$ målt i θ_0 . I ett år analysen ovenfor ble \hat{a} (Russland) overestimert med 36.5 % som figuren illustrerer. Y-aksen er begrenset til ± 25 og det er forøvrig 1 punkt som faller utenfor denne grensen. Fra 1985 er parameter a først underestimert og deretter kraftig overestimert frem til 2008. a (Norge) viser minimale skjevheter gjennom hele perioden hvor regresjonslinjen tilpasser punktene meget godt.

Spredningen på skjevheten til parameter b er større for russiske menn, men den gjennomsnittlige kurven er relativ lik. Forskjellen er at \hat{b} (Norge) er overestimert i perioden 1960-1970 og fra 2000 til 2009, mens \hat{b} (Russland) er henholdsvis overestimert og underestimert fra 1980 til 2000 og 2000 til 2009. Den gjennomsnittlige skjevheten til parameteren c er et speilbilde av parameter b . Periodene der \hat{b} er overestimert viser figuren at \hat{c} er underestimert. Årsaken til dette er å kompensere estimeringen slik at bootstrap-dødeligheten blir liggende nærmest mulig q_{θ_0} . Forøvrig er spredningen størst for russiske menn.

Ekstra usikkerhet

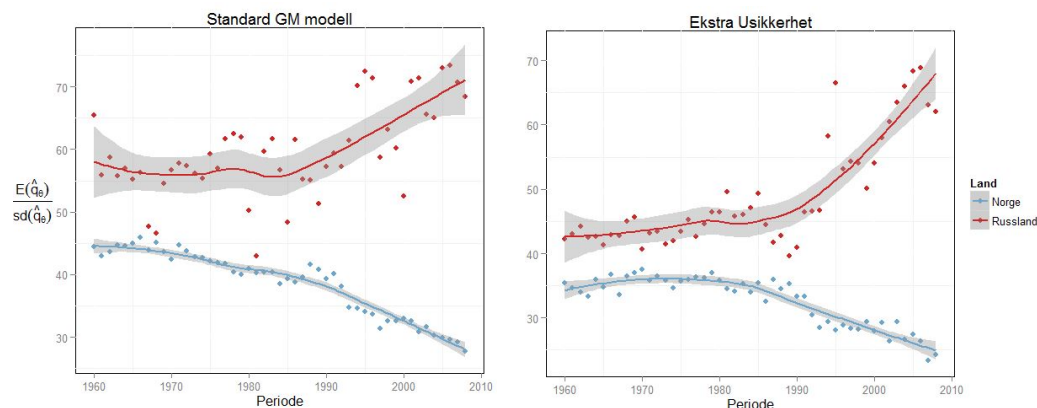
En økende γ i den beta-binomiske dødelighetsmodellen fører til et lavere standardavvik.



Figur 5.10: Viser utviklingen til γ for diverse land for perioden 1960 til 2009. Det er blitt brukt reelle datastørrelser.

Figur 5.10 viser utviklingen til γ i perioden 1960 - 2009. Resultatene er interessante. Det er stabil og jevn økning for landene Italia, Japan, Sverige og Norge. USA skiller seg ut etter 1990, mens γ for russiske menn knapt øker. Bakgrunnen for dette er i tråd med utviklingen i dødelighetstrenden. De landene som har hatt en markant reduksjon i dødelighet har en tilsvarende reduksjon i standardavviket. Det gir mening. Med utgangspunkt i figuren ovenfor er det enkelt å regne ut standardavviket for alder x i periode k . Det er naturlig at standardavviket er større for russiske menn siden dødeligheten er høyere.

5.2. AVVIK I PENSJONSEVALUERINGER



Figur 5.11: Viser forholdet mellom forventning og standardavvik for henholdsvis standard Gompertz-Makeham og Ekstra usikkerhet i perioden 1960 til 2009. $N = 1\,000\,000$

Figur 5.14 viser gjennomsnittlig forhold mellom dødelighet og standardavvik i perioden 1960 til 2008 med en eksponeringsstørrelse på en million mennesker. Forholdet er lavere i den beta-binomiske modellen og viser større usikkerhet en Gompertz-Makeham modellen. Trenden mellom modellene er identiske som viser tegn til stabilitet. Det er henholdsvis avtagende og økende trend for norske og russiske menn. Regersjonslinjen viser større svingninger for russiske menn som skyldes landets dødelighetsutvikling.

5.2 Avvik i pensjonsevalueringer

Begreper	Symbol	Antakelse	Begreper	Symbol	Antakelse
Tid per utbetaling	T	Årlig	Rente	r	3%
Pensjonsalder	x_r	65 år	Høyeste alder	x_e	110 år
Pensjonutbetaling	s	150 000 kroner	Pensjonspremie	π	ekvivalent

Tabell 5.6: Antakelser/Begreper for beregning av engangspremieberegninger

I denne seksjonen beregnes engangspremiefeil med antakelser hentet fra tabell 5.6 på en rekke land på bakgrunn fra feilanalysen i kapittel 3.

5.2.1 År 2009

	Russland	Norge	Usa	Japan	Italia	Sverige
$\Psi(g_{30})$	261 029	630 376	587 962	651 617	641 543	645 350
$\Psi(g_{50})$	573 508	1 168 285	1 111 070	1 209 525	1 185 165	1 191 509

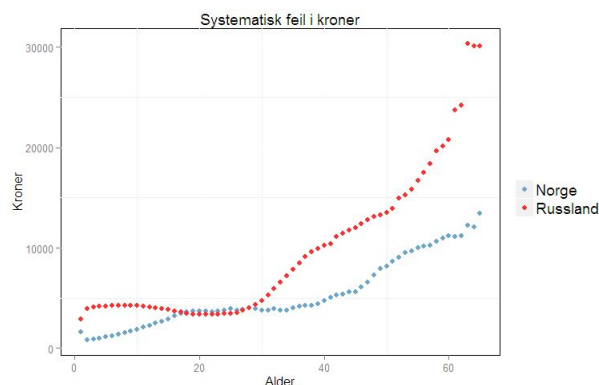
Tabell 5.7: Korrekt premie for alder 30 og 50 år i en rekke land

Tabell 5.7 viser kostnaden på en korrekt engangspremie for alder 30 og 50 år for å yte et forhåndsavtalt beløp hvert år i pensjonsalder. Det er gjort på grunnlag fra

5.2. AVVIK I PENSJONSEVALUERINGER

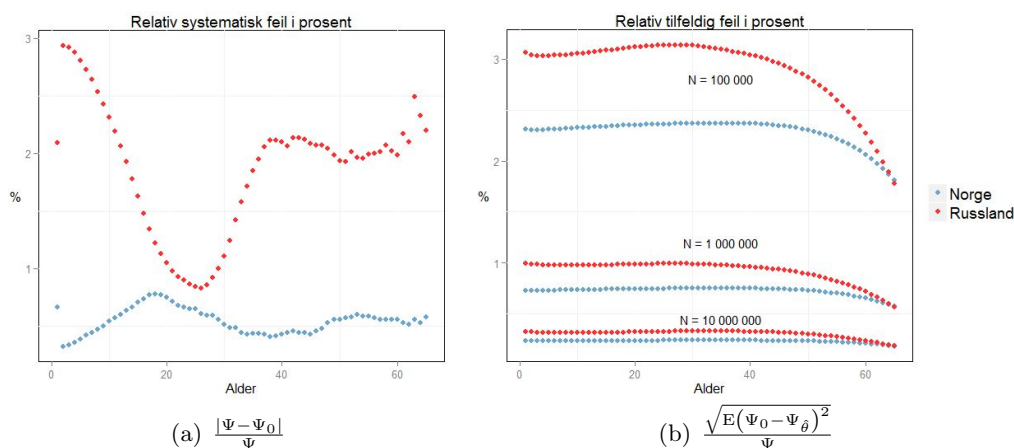
året 2009 for landene Russland, Norge, USA, Japan, Italia og Sverige. Det er tydelig forskjell mellom landene, og ikke uventet er prisen for russiske menn rimeligst. En russisk femtiåring betaler mindre premie enn det en trettiåring betaler i de resterende land for identiske vilkår. Tabellen viser hvor utslagsgivende dødeligheten virker inn på engangspremier. Dette er en god illustrasjon hvor viktig det er å bruke data fra riktig populasjon.

Figur 5.12 viser feil i engangspremiene for norske og russiske menn i alder 1-65 år. Alderen 30 til 65 år er fasen hvor flertallet av befolkningen er yrkesaktive og det



Figur 5.12: Figuren viser total og systematisk feil for norske og russiske menn fra 2009.

investeres i pensjon. Det er en viktig alder både for livselskapet og pensjonstaker. Figuren viser at systematisk feil er større hos russiske menn i denne alderen. I tillegg er russiske engangspremier langt lavere sammenlignet med norske menn i tilsvarende alder.



Figur 5.13: Figur a) viser relativ systematisk feil i prosent for norske og russiske menn i alderen 1 til 65 år. Figur b) relativ tilfeldig feil i prosent for tilsvarende land og alder. Det er åpenbart størst forskjell i modellfeilen.

Figur 5.13 a) viser relativ systematisk feil målt i forhold til korrekt engangspremie for norske og russiske menn. Relativt sett, er det stor forskjell mellom landene. Fra

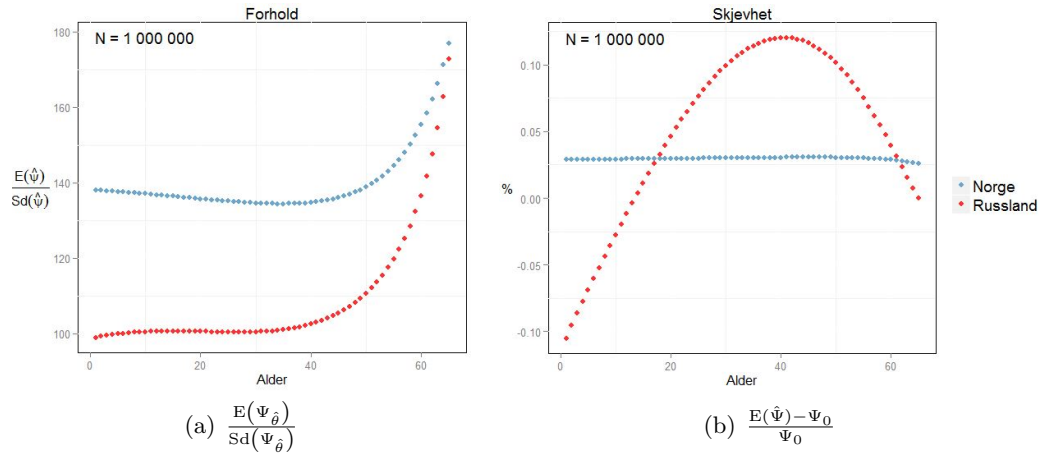
5.2. AVVIK I PENSJONSEVALUERINGER

alderen 30 til 65 år er relativ modellfeil på russiske menn godt over fire ganger så stor. Det er store svinginger for russiske menn som betyr at modellfeilen er i større grad aldersavhengig. Forøvrig er modellfeilen sett med norske øyne tilfredstillende med en gjennomsnittverdi på 0.62 % og et standardavvik på 0.11 %. Russiske menn har tilsvarende verdier på henholdsvis 2.44 % og 0.53 %. Se tabell 5.8.

Systematisk feil	Norge	Russland
$E_{x=1:65} \left(\frac{ \Psi - \Psi_0 }{\Psi} \right)$	0.62 %	2.44 %
$Sd_{x=1:65} \left(\frac{ \Psi - \Psi_0 }{\Psi} \right)$	0.11 %	0.53 %

Tabell 5.8: Viser gjennomsnitt og standardavvik til relativ systematisk feil for norske og russiske menn i alderen 1 til 65 år.

Figur 5.13 b) viser relativ tilfeldig feil i tilsvarende alder. Eksponeringsstørrelse på hundre tusen gir stor feil for begge land, men avtar ettersom porteføljestørrelsen øker. Forskjellen i den relative tilfeldige feilen mellom landene er mindre sammelignet med modellfeilen. Det er derfor modellfeilen som utgjør forskjellen, noe figuren tydelig illustrer.

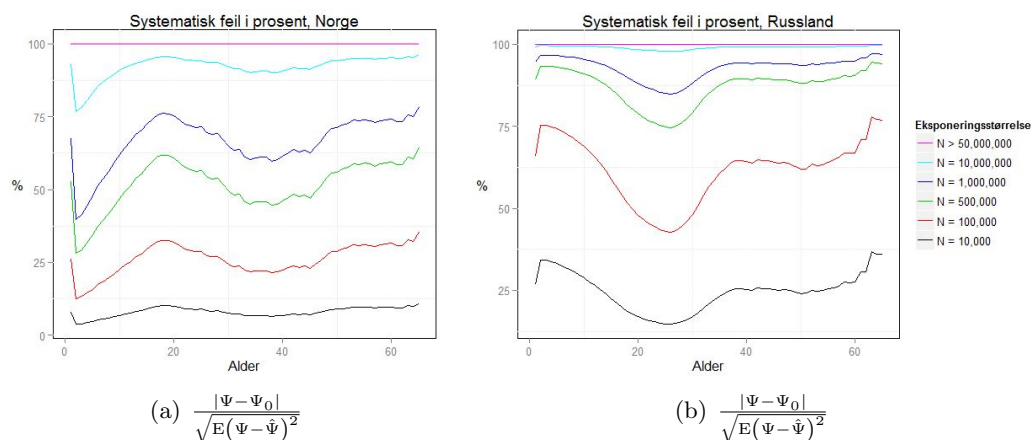


Figur 5.14: Figur a) viser forholdet mellom forventning og standardavvik. Usikkerhet er åpenbart lav. Figur b) viser skjevheten til engangspremien ved bootstrap. Det er større svinginger i russiske menn.

Figur 5.14 viser forholdstest og skjevhet for norske og russiske menn i alderen 1-65 år når pensjonsporteføljen løper på en million mennesker. Begge land har et forhold på godt over 100. Det betyr at bootstrap-simuleringene har minimale svingninger i forhold til forventet verdi, Ψ_0 . Høyre graf viser skjevhet til engangspremiene. Bootstrap-premiene til norske menn overestimert på rundt 0.05 %. $\hat{\Psi}$ (Russland) er mer variabel hvor det både er over og under-estimering i intervallet -0.10 til 0.10 %. Siden bootstrap-dødeligheten er overestimert i alderen 0 til 20 år, gir det rimeligere premie som igjen fører til underestimering. Tilfellet er motsatt i alderen 20 til 75. Sammenlignet med parameter-skjevhet fra forrige seksjon er skjevheten til engangspremiene relativt sett lavere. Se tabell 5.4.

5.2. AVVIK I PENSJONSEVALUERINGER

Figur 5.15 viser systematisk feil i prosent for norske og russiske menn i alderen 1 til 65 år. Siden avviket $|\Psi_0 - \hat{\Psi}|$ går mot 0, vil systematisk feil dominere pensjonsporteføljene når eksponeringsstørrelsen er tilstrekkelig stor. Figuren illustrerer nettopp det. Ikke uventet konvergerer systematisk feil i prosent raskere og domineres i større grad for russiske menn. Årsaken til det er at relativ systematisk feil for russiske menn er større. I tillegg konvergerer estimeringsfeilen mellom landene likt. Se figur 5.13. Følgelig vil modellfeilen dominere raskere. Russisk data er på ingen måte bedre egnet. Derimot, er det illustrert mindre feil og mer stabilitet i norsk dødelighetsdata under Gompertz-Makeham modellen.



Figur 5.15: Viser systematisk feil i prosent for en rekke forskjellige eksponeringsstørrelser for norske og russiske menn fra året 2009

Figuren viser et eksponeringsbehov på godt over ti millioner norske menn for å oppfylle en modellfeil på 100 %. Det er ikke realistisk ettersom befolkningen ligger på 5 million mennesker hvor av halvparten er kvinner. Det er blitt brukt en aldersprofil med toppunkt på 40 år og en $\gamma = 0.10$ som gir flest mulig i de viktige aldrene fra 20 til 60 år. En pensjonsportefølje på en million mennesker tilsvarer 45 623 førtiåringer, og henholdsvis 16 700 og 20 500 er tretti og fremti år.

2009	N = 1 000 000					
x	10 år	20 år	30 år	40 år	50 år	65 år
$\frac{\sqrt{E(\Psi - \hat{\Psi})^2}}{\Psi}$ (Nor)	0.89 %	1.06 %	0.93 %	0.90 %	0.99 %	0.88 %
$\frac{\sqrt{E(\Psi - \hat{\Psi})^2}}{\Psi}$ (Rus)	3.23 %	2.02 %	1.99 %	2.75 %	2.42 %	2.28 %

Tabell 5.9: Viser RRKF for norske og russiske menn i diverse aldre når pensjonsporteføljen løper på en million mennesker.

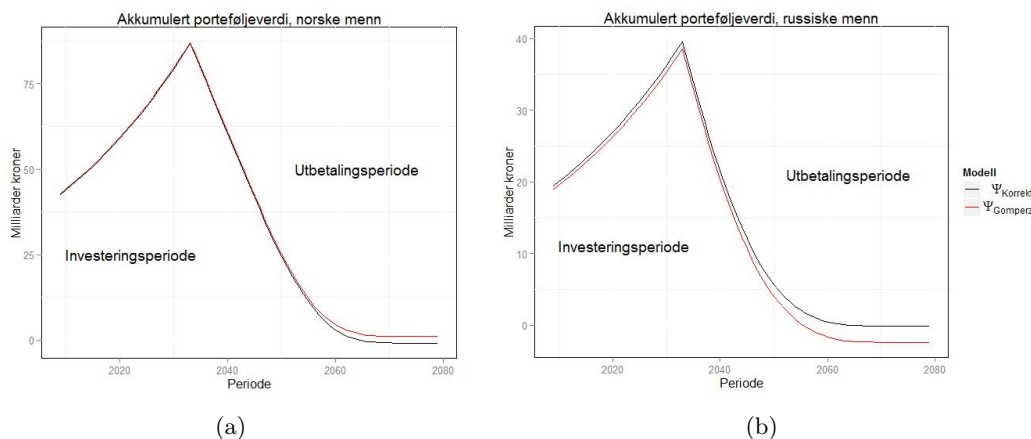
Hvis eksponeringsstørrelsen derimot er ti millioner tilsvarer det nesten en halv million førti-åringer. Det er mye målt opp mot befolkningen i Norge. Allikevel er det tilfredsstillende resultater ettersom systematisk feil er svært lav. En pensjonsportefølje på rundt 1 million mennesker gir i underkant av 1 % total relativ feil. Russiske menn krever langt mindre eksponeringsstørrelse for å oppnå tilsvarende systematisk domi-

5.2. AVVIK I PENSJONSEVALUERINGER

nans. Bakgrunnen er allerede diskutert. Modellfeilen dominerer ved rundt en million mennesker med tilsvarende aldersprofil som i Norge. Russland har en befolkning på godt over 143 millioner hvorav rundt 46 % er menn. En pensjonsportefølje på en million menn er realistisk for et russisk forsikringsselskap. En pensjonsportefølje på rundt en million mennesker vil gi omtrent 1.5 til 3.5 % relativ total feil avhengig av alder. Det er en langt større total feil sammenlignet med Norge. Forskjellen ligger åpenbart i modellfeilen. Se Tabell 5.9.

Modellfeil i et økonomisk eksperiment

Eksponeringsstørrelse på en million tilsvarende omtrent 45 000 førtiåringer. Anta nå at disse tegner engangspremie i et forsikringsselskap. Engangspremien er beregnet på bakgrunn fra parametersettet θ_0 . Verdien på porteføljen oppdateres hver periode frem til kontraktene dør ut. Se algoritme 4.2.4. Oppdateringen baserer seg på det som faktisk skjer slik at korrekt dødelighet g_x brukes.



Figur 5.16: Viser akkumulert porteføljeverdi for norske og russiske menn. Utgangspunktet er 45 000 førtiåringer. Merk y-aksene.

Figur 5.16 viser utviklingen til akkumulert porteføljeverdi. Det er større forskjell i porteføljen for russiske menn. Legg merke til at det er synlige avvik allerede i investeringsperioden. Tabell 5.10 viser utviklingen i utvalgte perioder. Det mest utslagsgivende er at porteføljen for russiske menn ender med et underskudd, mens motsatt er tilfellet for norske menn. Årsaken for det er at Gompertz-Makeham modellen under og overestimerer henholdsvis gjennomsnittlig dødelighet for norske og russiske menn. Tabell og figur illustrer det. Det tilfaller en fordel for et norsk forsikringsselskap som nesten har 2 milliarder kroner i *overskudd*. I Russland tilfaller det en fordel for den gjennomsnittlige forsikringstakeren.

5.2. AVVIK I PENSJONSEVALUERINGER

(a) Russland					
Russland	2035	2045	2055	2065	2080
Ψ_{Korrekt}	39.70	15.63	3.38	0.21	0
Ψ_{gompertz}	38.71	14.30	1.60	-2.05	-2.25
(b) Norge					
Norge	2035	2045	2055	2065	2080
Ψ_{Korrekt}	87.55	49.87	16.83	1.44	0
Ψ_{gompertz}	87.98	50.46	17.62	3.31	1.86

Tabell 5.10: Viser akkumulert porteføljeverdi for norske og russiske menn. Modellfeilen tilfaller en fordel for et norsk livselskap. Motsatt er tilfellet for russisk selskap.

Ugunstig utvalg

Seleksjonsproblemet er et velkjent problem for forsikringsselskaper. For eksempel i det mest ekstreme, et norsk forsikringsselskap må utelukkende ikke basere fremtidige premier, avsetninger og utbetalinger på russisk dødeligheter. Resultatene ovenfor illustrerer det tydelig! Denne oppgaven bruker dødelighetsdata på bakgrunn fra hele befolkningen, altså landsgjennomsnittet. Informasjonen i markedet er antatt symmetrisk, som innebærer at hvert menneske løper på lik dødelighetsrisiko. Det er i utgangspunktet naivt, og fører til store økonomiske problemer for selskapene.

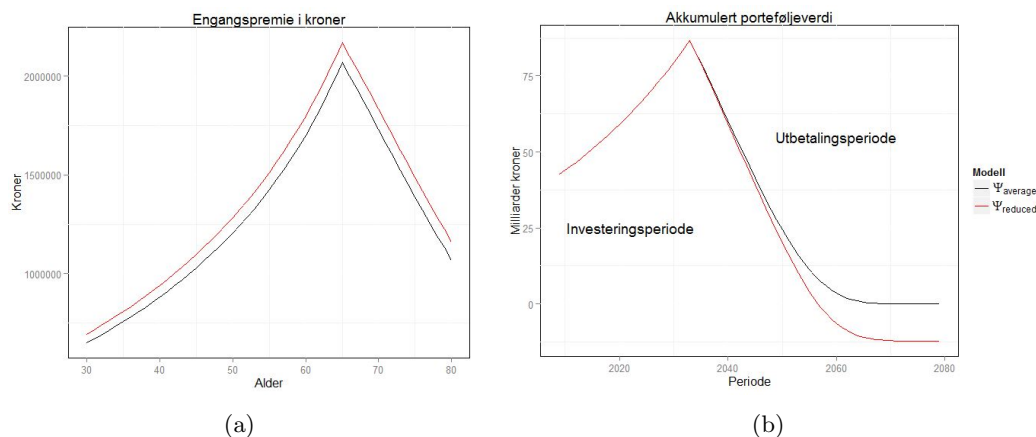
Kort fortalt oppstår problemet når mennesker med henholdsvis høy og lav dødelighetsrisiko må betale lik premie. Dette er urettferdig for menneskene med høy risiko som subsidierer de med lav risiko. I det lange løp fører det til en ujevn pensjonsportefølje hvor flertallet er representert fra gruppen med lav risiko. Hendvender leseren til [6] for ytterligere bakgrunn, og forslag til å minimalisere problemet. I situasjoner der det velges feil datagrunnlag hjelper det lite at modellen er nøyaktig. Nedenfor er det gjort et eksperiment.

Økonomisk eksperiment ved ulike levealdre

En rimelig antakelse er å anta at mennesker i en forsikringsportefølje har 10 - 15 % lavere dødelighet enn landsgjennomsnittet i hver alder[1]. Figur 5.17 a) viser engangspremiekurven med og uten(Ψ) reduksjon i dødelighet i alder 30 til 80 år fra 2009. Engangspremiene utgjør stor forskjell, og illustrer hvor nødvendig det er å ta hensyn til ulike levealdre i en befolkning.

Anta et selskap tegner 45 000 engangspremier ved inngangen av 2010 for norske menn i alder 40 år. Uten hensyn til dødelighetsreduksjon får selskapet inn $45\,000 \cdot \Psi_{40} = 41.4$ milliarder kroner, og investerer i obligasjoner med årlig rente 3 %. Beholdningen ved inngang til pensjonsalder tilsvare 86.68 milliarder kroner. Videre blir levealderen og verdien på porteføljen (med rentjusteringer) oppdatert hver periode frem til kontraktene *dør ut*. Figur 5.11 b) viser utviklingen til porteføljen ved **to scenarier**. Et scenario hvor faktisk dødelighet følger landsgjennomsnittet, og det andre hvor faktisk dødelighet er 15 % lavere i hver alder.

5.2. AVVIK I PENSJONSEVALUERINGER



Figur 5.17: a) viser engangspremiekostnad for norske menn, alder 30 til 80 år ved to scenarier. b) viser akkumulert porteføljeværdi med hensyn på engangspremiene i a)

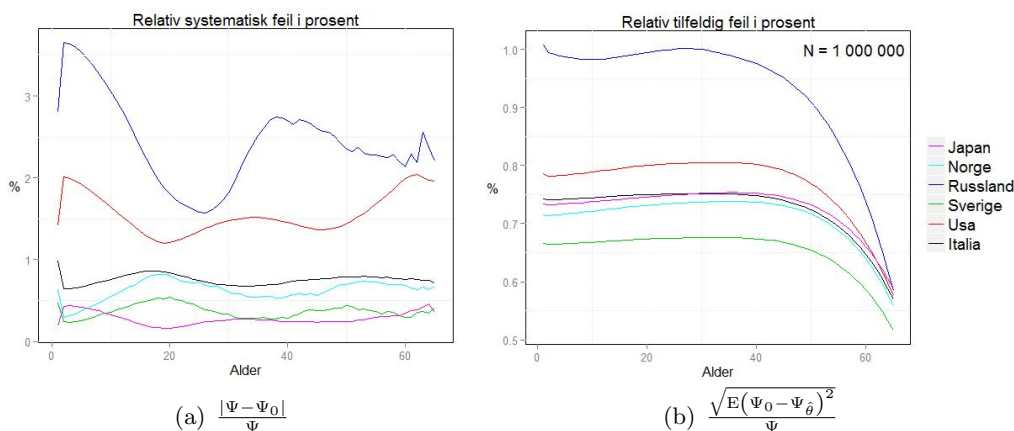
Tabell 5.11 viser portefølje-utviklingen i utvalgte perioder. Porteføljen til scenariet Ψ_{reduced} ender med et underskudd på over 12 milliarder kroner. Det er mye penger, og dessuten er den årlige pensjonsutbetalingen kun 150 000 kroner. Det vil muligens være større beløp i virkeligheten. Eksperimentet illustrer hvor viktig det er å ta utgangspunktet i riktig populasjon. Feilen mellom korrekt og parametrisert portefølje⁵ fra forrige avsnitt utgjør kun 15 % av feilen mellom korrekt og redusert portefølje. Det betyr hvis populasjonen er feil hjelper det lite at dødelighetsmodellen er nøyaktig!

Modell	2035	2045	2055	2065	2080
Ψ_{average}	86.66	49.87	16.83	1.44	0
Ψ_{reduced}	86.66	47.29	10.06	-9.72	-12.28

Tabell 5.11: Akkumulert porteføljeværdi, i milliarder kroner, for en pensjonskasse med 10 000 norske menn, alder 40 år.

⁵Porteføljefeilen mellom Ψ og Ψ_0 , altså modellfeilen

Italia, Japan, USA og Sverige



Figur 5.18: Figur a) viser relativ systematisk feil i prosent for en rekke land i alderen 1 til 65 år. Figur b) relativ tilfeldig feil i prosent for tilsvarende land og alder. Russland, og delvis USA skiller seg tydelig ut.

Figur 5.18 a) illustrerer relativ systematisk feil for landene året 2009. Modellfeilen er åpenbar størst for russisk menn, etterfulgt av de amerikanske. Deretter følger resterende land tett etter hverandre. Kurvene til de nordiske landene beveger seg tilsynelatende likt, mens japanske menn gir lavest modellfeil. En interessant sammenligning er å se på forventet levealder og relativ modellfeil. Tabell 5.12 viser tydelig en sammenheng hvor lavere forventet levealder gir høyere modellfeil⁶. Årsaken er muligens knyttet til landets dødelighetsprofil. Høy dødelighet gir manglende data for høye aldre. Det kan igjen føre til større avvik i modellen.

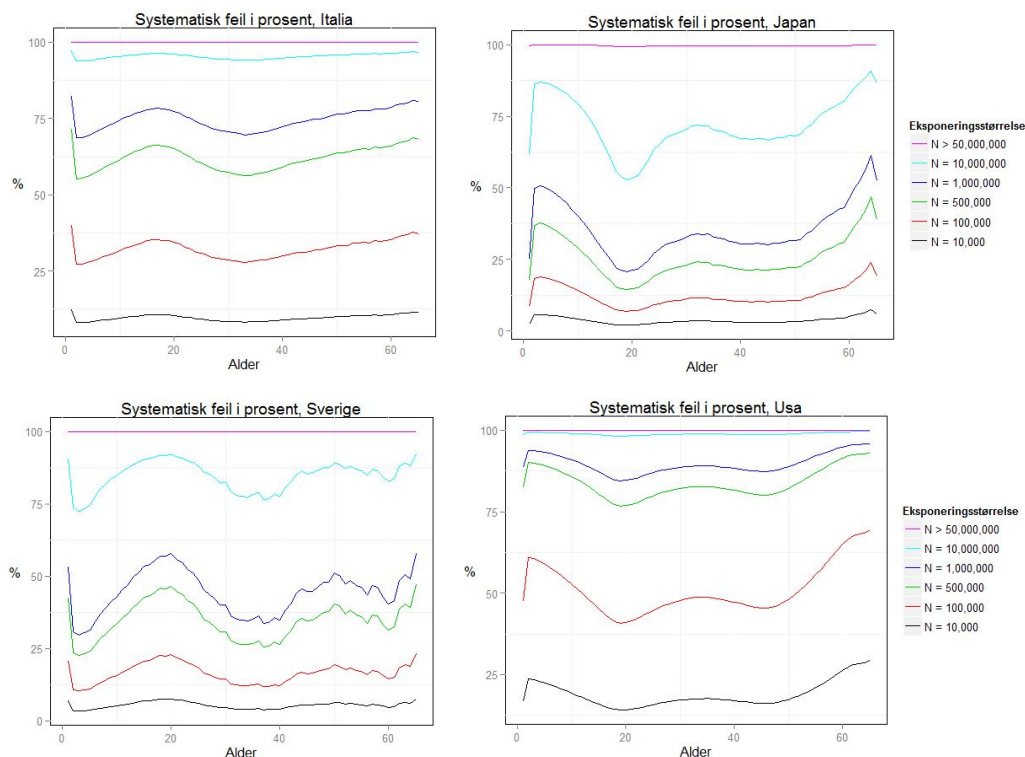
År 2009	Japan	Sverige	Norge	Italia	Usa	Russland
$E\left(\frac{ \Psi - \Psi_0 }{\Psi}\right)$	0.27 %	0.36 %	0.62 %	0.75 %	1.55 %	2.43 %
Forventet levalder	79.6 år	79.3 år	78.6 år	79.1 år	76.3 år	61.7 år

Tabell 5.12: Viser gjennomsnittelig relativ systematisk feil for en rekke land når pensjonsporteføljen løper på en million mennesker. Det er en tydelig negativ sammenheng mellom RSF og forventet levealder.

Figur 5.18 b) viser relativ tilfeldig feil når pensjonsporteføljen løper på en million mennesker. Russiske menn gir størst utslag. Allikevel er forholdet langt lavere til resterende land, sammenlignet med modellfeilen. Merk y-aksene. Figur 5.19 viser systematisk feil i prosent for landene Italia, Japan, USA og Sverige på dødelighetsdata fra 2009. Landene viser åpenbart flere likheter med Norge enn Russland. Japan og Sverige konvergerer noe tregere, mens Italia ligger på omtrent likt nivå. USA skiller seg noe ut, og er helt klart landet som har likhetstrekk med Russland. Bakgrunnen for det er i tråd med landets dødelighetsprofil.

⁶Gjennomsnittelig relativ systematisk feil

5.2. AVVIK I PENSJONSEVALUERINGER



Figur 5.19: Viser systematisk feil i prosent for en rekke forskjellige eksponeringsstørrelser for Italia, Japan, Sverige og Usa.

5.2.2 Periode 1960 til 2008

Tabell 5.13 viser korrekt engangspremie som må betales inn i alder 30 og 40 år i periodene 1960, 1970 og opp til 2008 for Norge og Russland. Ikke overraskende er trendene motgående i perioden 1970 til 2000. Det er en nedgang i engangspremiene for norske menn i perioden 1960-1970, men skyter fart frem mot 2008.

	k	1960	1970	1985	2000	2008
$\Psi_{30,k}$	Norge	483 766	445 545	480 598	549 814	621 741
	Russland	344 481	315 697	277 765	240 485	261 028
$\Psi_{40,k}$	Norge	660 288	608 140	653 714	747 140	842 611
	Russland	483 353	447 696	395 395	346 764	378 248

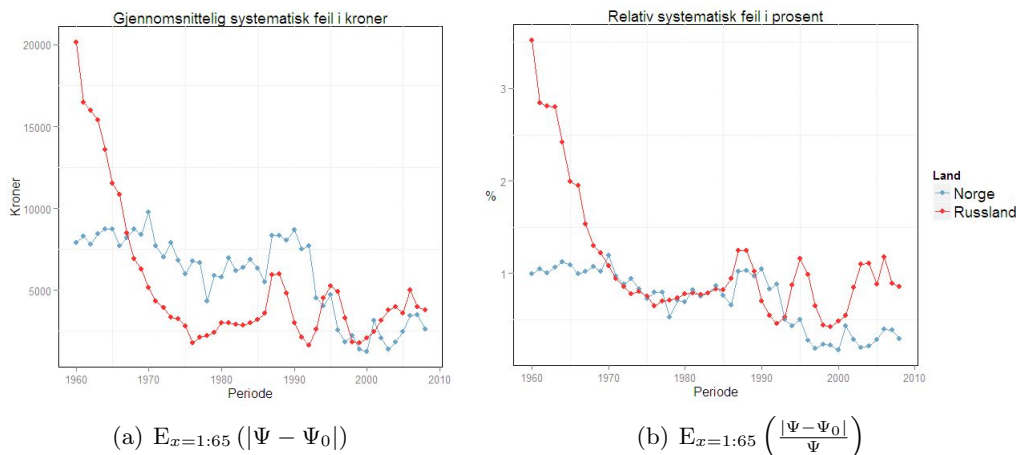
Tabell 5.13: Korrekt premie for alder 40 år i en rekke perioder for Norge og Russland

Figur 5.20 a) viser gjennomsnittlig⁷ systematisk feil i kroner fra perioden 1960 til 2008. Systematisk feil(Norge) har en nedadgående trend samtidig som det er vist at engangspremiene har økt i tilsvarende periode. Det betyr at Gompertz-Makeham modellen muligens beskriver de senere periodene bedre. Russiske menn har tilsva-

⁷ Alderen 1 til 65 år

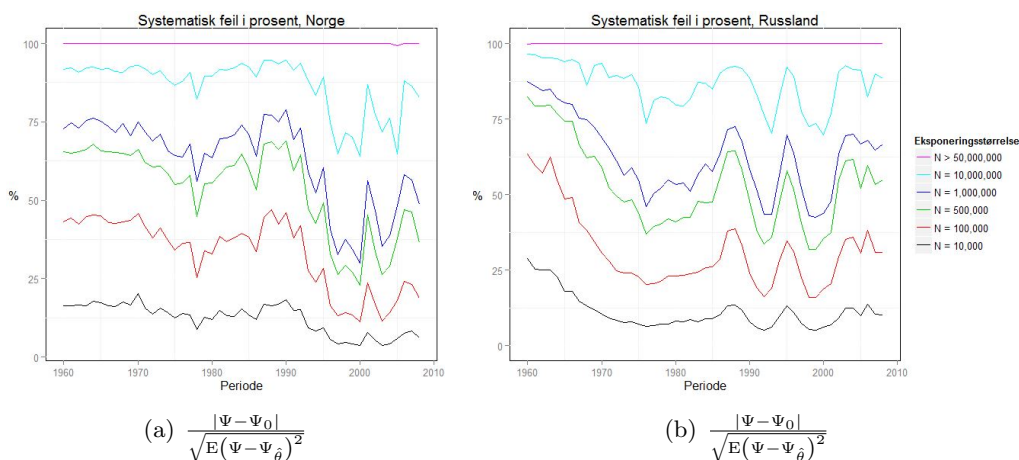
5.2. AVVIK I PENSJONSEVALUERINGER

rende trend. Eneste forskjell er at russiske engangspremier har økt slik at det er vanskelig å si noe håndfast.



Figur 5.20: Figur a) viser systematisk feil for norske og russiske menn i perioden 1960-2009 i alderen 1 til 65 år. Figur b) viser gjennomsnittlig relativ systematisk feil for norske og russiske menn i alder 1 til 65 år. Perioden løper fra 1960 til 2008. Trendene er avtagende.

Figur 5.20 b) viser gjennomsnittlig relativ systematisk feil i tilsvarende periode. Den relative modellfeilen er tydelig avtagende i perioden. Hvor norske menn har en jevn og stabil nedgang er russiske menn i større grad varierende. Spesielt fra 1960 til 1970 er det tilsynelatende stor feil med likhetstrekk fra året 2009. Riktignok er resultatene mer stabile fra 1970 til 1995. Det er også enkelte år hvor modellfeilen er lavere enn for norske menn. Fra 1996 og frem til i 2008 er det igjen større feil og svingninger i russisk dødelighetsdata



Figur 5.21: Figuren viser gjennomsnittlig systematisk feil i % for norske og russiske menn i perioden 1960-2009. Grafen viser den gjennomsnittlige feilen i alder 1 til 65 år.

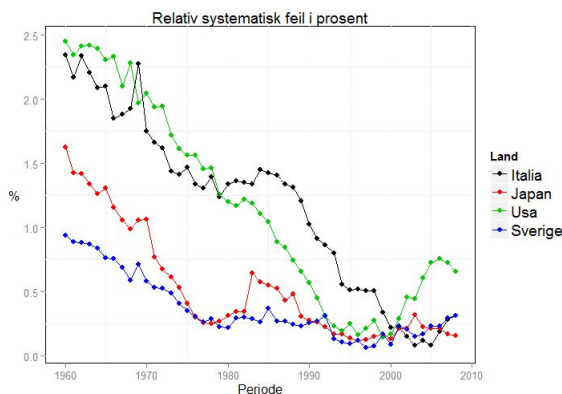
Allikevel er feilene i perioden langt lavere enn i 2009, vist ovenfor. Maksimal levealder er satt til 105 år grunnet manglende data fra HMD. Dette reduserer muligens

5.2. AVVIK I PENSJONSEVALUERINGER

modellfeilen for russiske menn ettersom det allerede er svært få som oppnår alder 110 år. Det var alderen som ble brukt i 2009. Relativ feil for norske menn halveres fra 1 til 0.5 %, og indikerer tydelig at Gompertz-Makeham modellen beskriver dødeligheten bedre med tiden. Det er nærliggende å trekke paralleller til den økende levealderen i tilsvarende periode.

Figur 5.21 viser gjennomsnittelig systematisk feil i prosent for norske og russiske menn. Perioden 1960 til 1970 og årene rundt 2000-tallet vokser modellfeilen i prosent raskere med eksponeringsstørrelsen for russiske menn. Fra 1970 til 1995 er modellfeilen tilsynelatende lik. Det er en svak nedadgående trend for norske menn som styrker argumentet for at Gompertz-Makeham modellen beskrives bedre med tiden. Trenden til russiske menn er noe mer variabel. For norske menn i perioden 1960 til 1990 utgjør modellfeilen i engangspremiene omtrent 75 % når pensjonsporteføljen er eksponert mot en million mennesker. I de senere årene faller modellfeilen ned omtrent til 50 %, som stemmer godt overens med resultatet fra 2009. Totalt sett fører det til små avvik. Kort oppsummert er det totalt mer feil og variasjon i pensjonsporteføljene til russiske menn. Periodene 1960 til 1970 og 2004 frem til idag er det stor forskjell, mens 1970 til 2000 er mer lik.

Italia, Japan, USA og Sverige

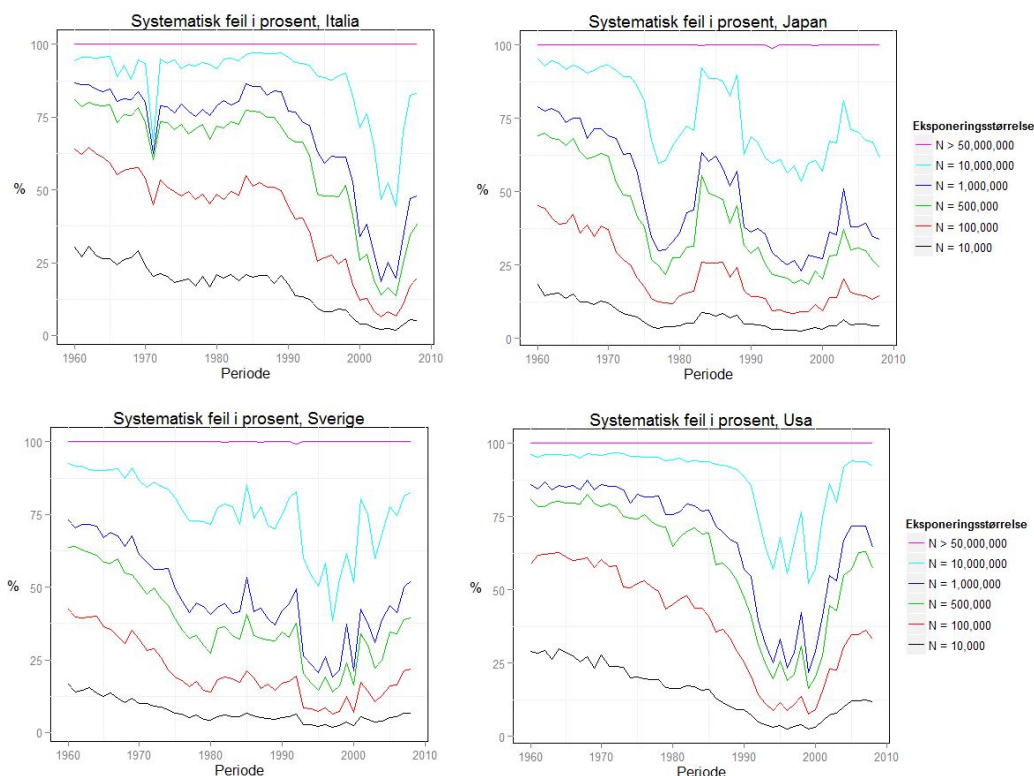


Figur 5.22: Viser relativ systematisk feil i prosent for Italia, Usa, Japan og Sverige. Det er en tydelig nedadgående trend med tiden.

Figur 5.22 viser relativ systematisk feil i prosent i en pensjonsportefølje for Italia, Japan, Usa og Sverige. Utviklingen er tilsvarende med det som er vist tidligere. Relativ modellfeil reduseres med tiden i alle tilfellene. Japan og Sverige har store likhetstrekk med Norge. Nedgangen i modellfeilen skjer i en periode hvor levealderen har økt kraftig. Om det har en sammenheng eller ikke er vanskelig å si, og ligger utenfor oppgavens omfang. Lengre levetid og lavere dødelighet gir bedre datagrunnlag.

Figur 5.23 viser systematisk feil i prosent i forhold til RKF. Det er opplagt likheter sammenlignet med norske menn. Se figur 5.21 a). Sverige og Japan har en fallende vekst med årene som betyr at modellfeilen blir mindre. Det er diskutert tidligere.

5.3. ASYMPTOTIKK MOT BOOTSTRAPPING



Figur 5.23: Viser gjennomsnittlig systematisk feil i prosent i alder 1 til 65 år for en rekke forskjellige eksponeringsstørrelser for Italia, Japan, Sverige og Usa.

5.3 Asymptotikk mot bootstrapping

Hittil er det blitt brukt bootstrap-simuleringer til å regne på den tilfeldige feilen. Istedenfor bootstrap kan det brukes asymptotiske utregninger. Det er interessant å sjekke hvor godt den asymptotiske teorien stemmer med bootstrap simuleringene fra forrige seksjon. Uansett, bootstrap viser seg å være klart den raskeste metoden, og gir meget bra resultater. Nedenfor er det gjort beregninger for perioden 1960 til 2009.

Tabell 5.14 og 5.15 viser standardavvikene til parameterene a, b og c mellom asymptotiske utregninger og bootstrap simuleringer for norske og russiske menn fra året 2009. De asymptotiske utregningene med hensyn på norske menn ligger opplagt nærmere bootstrap estimatene, spesielt parameterene a og b. De prosentmessige avvikene ved eksponeringsstørrelse på 50 000 og 500 000 norske menn er veldig nøyaktige. Det er stor forskjell sammenlignet med russiske menn.

Ifølge teorien i kapittel 3 vil standardavvikene ved begge metodene konvergere mot 0 slik at estimatene konvergerer mot θ_0 . Tabellen ovenfor gir indikasjon at asymptotiske utregninger med hensyn på russiske menn konvergerer tregere og vil følgelig trenge enn høyere eksponeringsstørrelse. Årsaken til dette er uklar, men i samtråd med det som er vist i seksjonene ovenfor er det grunnlag å tro det er mer usikkerhet rundt

5.4. OPPSUMMERING

Norge						
$m_b = 150$	N = 5 000		N = 50 000		N = 500 000	
	Bootstrap	As.	Bootstrap	As.	Bootstrap	As.
S_a	0.000574	0.000644	0.000204	0.000208	0.000065	0.000064
S_b	0.000018	0.000014	0.00000518	0.0000046	0.0000014	0.0000015
S_c	0.017	0.015	0.00492	0.00491	0.00154	0.00170

Tabell 5.14: Det er kjørt 500 bootstrap simuleringer.

Russland						
$m_b = 150$	N = 5 000		N = 50 000		N = 500 000	
	Bootstrap	As.	Bootstrap	As.	Bootstrap	As.
S_a	0.00190	0.00278	0.000704	0.000956	0.000265	0.000440
S_b	0.00032	0.00048	0.000118	0.000165	0.0000472	0.0000721
S_c	0.011	0.010	0.00283	0.00345	0.00102	0.00151

Tabell 5.15: Det er kjørt 500 bootstrap simuleringer.

modellen når russisk dødelighetsdata blir brukt.

5.4 Oppsummering

Hensikten med oppgaven var å studere estimering, modell og seleksjonsfeil i dødelighetsmodellering tilknyttet livsforsikring. Det ble utført et detaljert studie fra 2009. Dødelighetsnivået mellom landene er vidt forskjellig, og resulterer i langt dyrere engangspremie for norske menn. En norsk førtiåring må i 2009 betale 140 % mer i premie enn tilsvarende russisk førtiåring for like vilkår. Forskjellen er ekstrem, og et norsk forsikringselskap kan åpenbart ikke bruke russisk dødelighetsdata til å beregne pensjon.

Modellen er nøyaktig, spesielt på bakgrunn fra norsk dødelighetsdata. Modellfeil for norske menn utgjør kun 0.5 % i forhold til korrekt engangspremie. Det er i de fleste tilfeller kun en fjerdedel sammenlignet med russiske menn. Det er åpenbart den systematiske feilen som utgjør forskjell mellom norske og russiske menn. Altså det tilfeldige avviket $E(\hat{\psi}) - \psi_0$ går mot 0 omtrent ved tilsvarende eksponeringsstørrelse. Dessuten, når porteføljen løper på en million mennesker eller mer, utgjør tilfeldig feil i forhold til korrekt mindre enn 1 %.

Systematisk feil i prosent for russiske menn dominerer raskere med eksponeringsstørrelse. Modellfeilen utgjør opp mot 100 prosent når eksponeringen på porteføljen er en million. Total feil for russiske menn er størst både i kroner og relativt sett. Det skyldes modellfeil, og ikke estimeringsfeil. I tillegg er det store avvik fra landene Japan, Sverige, Italia og delvis USA. Det bekrefter at russisk dødelighetsdata skiller seg ut. Se figur 5.19.

Portefølje-eksperimentet mellom norske og russiske menn viste et interessant resultat. Modellfeilen sørget for at porteføljen endte med et overskudd på et par milliarder kroner⁸ for norske menn. Motsatt var tilfellet for russiske menn i tilsvarende eksperiment. Årsaken skyldes at dødeligheten er under- og overestimert henholdsvis for norske og russiske menn. Seleksjon-eksperimentet belyste hvor viktig det er å bruke riktig data. Hvis populasjonssgrunnlaget er feil hjelper det lite at modellen er nøyaktig. Se Figur 5.17 og tabell 5.11.

Den store forskjellen mellom Norge og Russland fra 2009 er ikke like tydelig i perioden 1960 til 2008. Bortsett fra perioden 1960 til 1970, og årene rundt 2000-tallet er modellen nøyaktig. Se figur 5.6. Det er viktig å understreke at studiet i denne perioden er langt mindre detaljert sammenlignet med året 2009. Den største forskjellen mellom landene ligger i stabiliteten. Årsaken skyldes turbulensen i russisk dødelighet med svært varierende svingninger. Dette kommer tydelig frem i parameter-analysen. Se figur 5.7, 5.8 og 5.9. Relativ total feil i engangspremie er avtagende i perioden 1960 til 2008. Trenden til tilfeldig feil er stabil. Det betyr at det er modellfeilen som reduserer total feil. Se figur 5.12.

Totalt sett beskriver Gompertz-Makeham modellen dødelighet godt i perioden 1960-2009 for Norge, Japan, Italia og Sverige. Resultatene viser klare bevis for at modellen har problem med dødelighetsdata for russiske menn i tilsvarende periode. Årsaken relateres til høy dødelighet og en påfølgende nedgang i levealder. Som en avsluttende del av kapitlet ble det vist at asymptotisk teori på modellen virker. Allikevel er bootstrap-simuleringer tidsmessig mye mer effektivt og derfor foretrukket ovenfor asymptotsk matematikk.

⁸45 000 førtiåringer med pensjonsutbetaling på 150 000 hvert år i pensjonsalder.

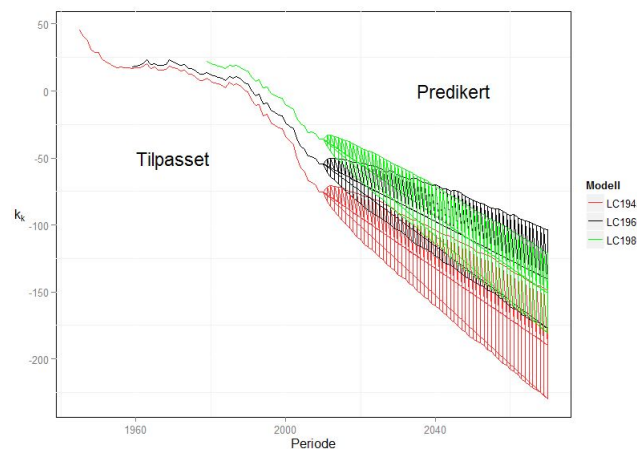
Kapittel 6

Lee-Carter

I dette kapitlet utføres numeriske anvendelser av Lee-Carter modellen med utgangspunkt fra norsk dødelighetsdata. Det brukes forskjellig grunnlagsperioder fra 1945 frem til 2009. I seksjonene som følger brukes notasjonen **LC1945**, **LC1960** og **LC1980**. Det er Lee-Carter modellen med utgangspunkt i henholdsvis periodene 1945-, 1960- og 1980-2009. Beregningene gjøres 60 år frem i tid, altså frem til 2070.

Seksjon 6.1 viser sensitiviteten i modellen fra diverse historiske grunnlagsperioder. Seksjon 6.2 tester Lee-Carter modellen på allerede kjent dødelighet. Det gjøres ved på bruke backtesting teorien(Dowd, 2008). Seksjon 6.3 viser hvordan lengre levetid kan påvirke engangspremie og pensjonsportefølje i regi av et økonomisk eksperiment.

6.1 Sensitivitet

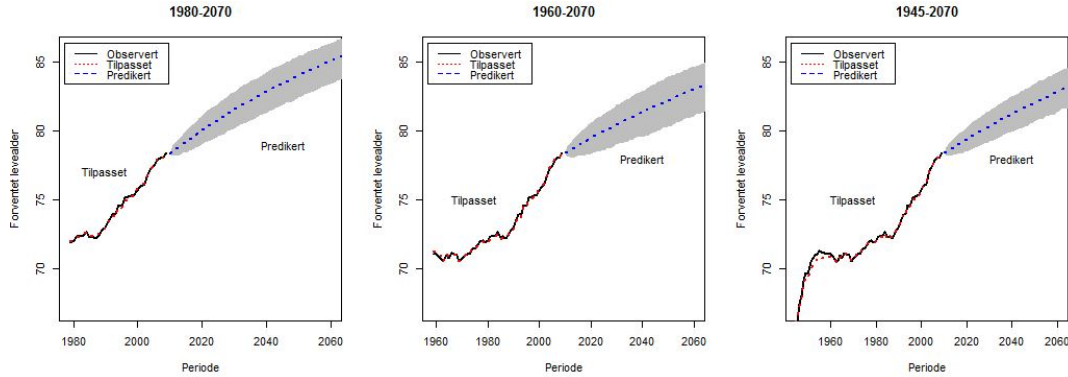


Figur 6.1: Viser utviklen til tidsrekken k_k for LC1945, LC1960 og LC1980. Det er større usikkerhet i sistnevnte sammenlignet med de andre periodene.

6.1. SENSITIVITET

Figur 6.1 viser historisk dødelighetstrend k_k og utviklingen til fremtidig dødelighetstrend k_{k+K} for LC1945, LC1960 og LC1980. Videre er det plottet henholdsvis 97.5 og 2.5 % persentiler som illustrer usikkerheten i henhold til hver modell. Konfidensintervallet mellom LC1960 og LC1945 er større enn LC1980. Det betyr at det er mindre varians og usikkerhet i sistnevnte modell. Årsaken til det følger fra datagrunnlaget der veksten i perioden 1980 til 2010 er mer moderat sammenlignet periodene 1945- og 1960-2010.

Figur 6.2 viser forventet levealder i modellene frem til 2060. Det er stor forskjell mellom LC1980 og de to andre modellene. Tabell 6.1 viser et utdrag fra figuren i periodene 2010, 2030 og 2060. Det er opplagte likheter mellom LC1960 og LC1945. LC1980 har høyest forventet levealder samtidig som usikkerheten er lavest. Kraftig dødelighetsreduksjon kombinert med moderat vekst i forventet levealder er naturlige årsaker for resultatene i modellen LC1980 ovenfor.



Figur 6.2: Viser forventet levealder i 2010-2060 med utgangspunkt i diverse baseperioder. LC1980 predikerer høyere forventet levetid. Det er mindre avvik mellom perioden 1945 og 1960 enn 1980.

I den originale Lee-Carter(1992)[17] artikkelen er det skrevet at modellen er stabil når datagrunnlaget er fra 30 til 90 år tilbake i tid. Her er det blitt brukt henholdsvis 30, 50 og 65 år tilbake i tid. Det betyr at 30 år muligens er litt snaut siden avviket til de resterende periodene er for stort.

	2010		2030		2070		
	$e_{0,2010}$	$e_{0,2030}$	2.5 %	97.5 %	$e_{0,2070}$	2.5 %	97.5 %
LC1980	78.4	81.5	80.1	82.8	86.0	84.4	87.2
LC1960	78.4	80.4	79.0	81.8	83.7	81.6	85.3
LC1945	78.4	80.3	79.1	81.4	83.6	81.8	85.0

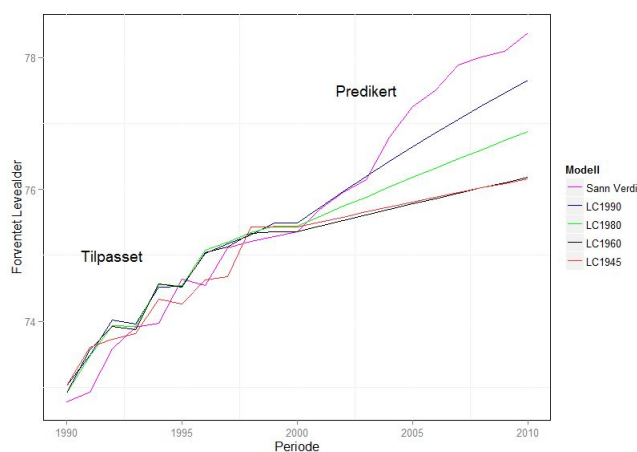
Tabell 6.1: Forventet levealder i fremtiden for LC1980, LC1960 og LC1945

Det er forventet en økning i levealder på 7.6 år fra 2010 til 2060 under LC1980. Gjennomsnittlig tilsvarer det en økning på 1.3 år for hvert tiår. LC1960 og LC45 har forventet økning på rundt 5.2 år i tilsvarende periode. For hvert tiende år som går øker levealderen med ett år. Det er stor forskjell mellom modellene. Perioden 1980

til 2010 er sannsynligvis ikke nok. Se [16].

6.2 Backtesting

I denne seksjonen utføres backtesting av allerede kjent dødelighet. I tillegg til periodene 1945-, 1960-, 1980-1999 er det også brukt periode 1990-1999, med navn LC1990. Figur 6.3 viser tilpasset forventet levealder frem til 1999, og predikert forventet levetid fra 2000 til 2009. Forskjellen er stor, både mellom Lee-Carter metodene og i forhold til den faktiske levealderen fra 2000 til 2009. LC1990 er den eneste som er i nærheten av den sanne levealderen, og LC1945 og LC1960 er veldig langt unna. Det er muligens flere grunner til dette. For det første var det en ekstrem økning på levealderen til norske menn fra 2000 til 2009 på 3 år. Det er tre ganger større enn tidligere vekst.



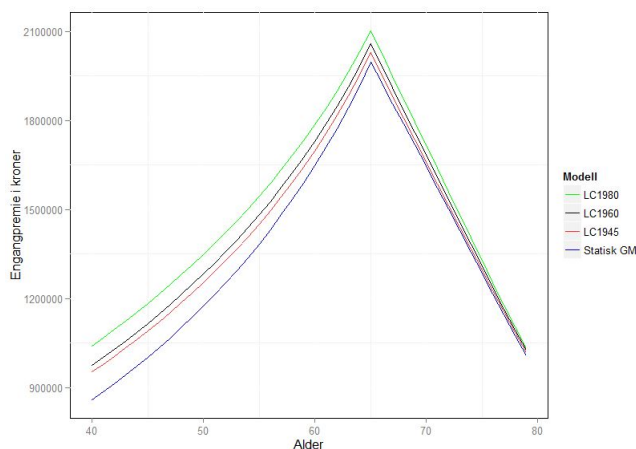
Figur 6.3: Viser backtesting for perioden 2000-2010 med hensyn på diverse basisperioder.

For det andre, LC1990 er kun modellert på bakgrunn av tiårsperioden 1990-1999. Det er ikke tilstrekkelig. I tillegg er det vært en tilsvarende kraftig vekst i denne perioden slik at det bare er naturlig at LC1990 ligger nærmest sann levealder. Videre er LC1945 og LC1960 igjen identiske, som viser på nytt stabiliteten til modellen ved å bruke lengre bakgrunn.

6.3 Pensjonsevalueringer med lengre levetid

Figur 6.4 viser kostnaden til en engangspremie i alder 40 til 80 år for modellene LC1945, LC1960 og LC1980. I tillegg er det lagt til et plott fra Gompertz-Makeham (GM statistisk) modellen fra forrige seksjon. Forskjellen er stor, spesielt kostnaden mellom GM-statisk og LC1980. Premien er omtrent lik for LC1945 og LC1960, som ikke overraskende avviker fra LC1980 der premien er betraktelig dyrere.

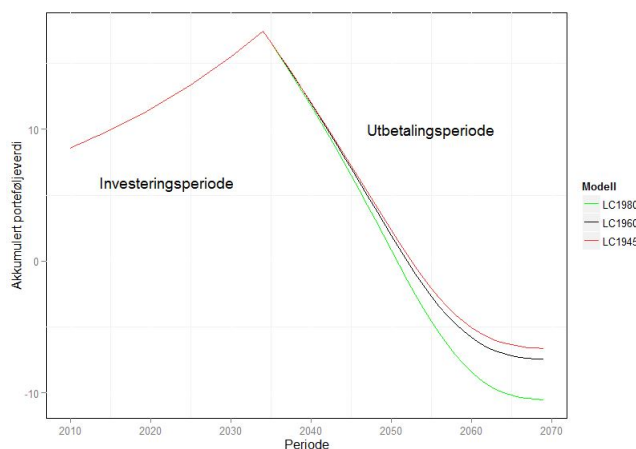
6.3. PENSJONSEVALUERINGER MED LENGRE LEVETID



Figur 6.4: Viser dynamisk engangspremier med hensyn på diverse basisperioder med Lee-Carter modellen. Videre er det plottet inn en statisk Gompertz-Makeham modell.

Økonomisk eksperiment ved lengre levetid

Et selskap tegner 10 000 engangspoliser ved inngangen av 2010 for norske menn i alder 40 år. Uten hensyn til dødelighetsreduksjon får selskapet inn $10\,000 \cdot 858\,848 = 8.58$ milliarder kroner, og investerer i obligasjoner med årlig rente 3 %. Det vil gi selskapet $8.58 \cdot 1.03^{25} = 17.45$ milliarder kroner¹ i inngangen til pensjonstiden. Videre antas det at fremtidig levetid følger LC1945, LC1960 og LC1980 slik at de faktiske dødelighetene blir modellert dynamisk. Videre blir levealderen og verdien på porteføljen (med rentejusteringer) oppdatert hver periode frem til kontrakten *dør ut*



Figur 6.5: Viser dynamisk engangspremier med hensyn på diverse basisperioder med Lee-Carter modellen. Videre er det plottet inn en statisk Gompertz-Makeham modell.

Figur 6.5 viser akkumulert portefølje-utvikling fra 2010 til 2055. I investeringspe-

¹Beløpet avviker fra GM-beløpet i seksjon 5.1 fordi her brukes det maksimum alder 100 år grunnet manglende data.

Modell	2035	2045	2055	2065	2070
LC1945	17.45	8.18	-1.24	-6.17	-6.62
LC1960	17.45	8.02	-1.85	-7.02	-7.47
LC1980	17.45	7.59	-3.52	-9.89	-10.45

Tabell 6.2: Akkumulert porteføljeverdi, i milliarder kroner, for en pensjonskasse med 10 000 norske menn, alder 40 år.

rioden vokser porteføljeverdien eksponentialt. Utbetalingsperioden starter i 2035. Ingen av porteføljene har nok kapital til å betale ut sine forpliktelser. Tabell 6.2 viser en oversikt fra hvert tiende år. LC1980 porteføljen er i underskudd allerede år 2050. Forsikringsselskapet sitter igjen med et tap på 10.45 milliarder kroner² hvis dødelighetsutviklingen følger LC1980 modellen. Det er ekstremt mye penger. Dessuten er porteføljestørrelsen liten samtidig som pensjonsutbetalingen er relativ lav. I virkeligheten kunne dette beløpet vært langt større. Eksperimentet viser også de økonomiske forskjellene ved å bruke forskjellige datagrunnlag. LC1980 avviker fra LC1945 og LC1960. Det stemmer overens med tidligere resultater!

6.4 Oppsummering

Som en avsluttende del av oppgaven ble det illustrert noen eksperimenter med Lee-Carter modellen. Modellen ble testet på norsk dødelighetsdata i forskjellige perioder fra 1960 frem til 2009. I den originale Lee-Carter(1992)[17] artikkelen er det skrevet at modellen er stabil hvis det brukes minst 30 år med historisk dødelighetsdata. Tabell 6.1 viser akkurat det. Periodene med horisont på 45 og 60 år gir omtrent lik predikering, mens perioden 1980 frem til 2009 skiller seg ut. Norske menn i 2060 forventes å leve til de er rundt 83 år. Det er en økning på nesten 5 år fra idag, altså ett år lenger for hvert tiår som går. Kostnaden av lengre levetid er stor. I eksperimentet i seksjon 6.3 forårsaket en pensjonsportefølje på 10 000 førtiåringer et underskudd i intervallet 6 til 12 milliarder kroner hvis det ikke tas hensyn til dødelighetsreduksjon. Det ble kun brukt en pensjonsutbetaling på 150 000 kroner. I virkeligheten er utbetalingen langt større slik at ytterligere tap er forventet med en tilsvarende porteføljestørrelse.

²Ikke tatt hensyn til kostnader i forbindelse med gjeld, i tillegg er maksimum levealder kun 100 år

Kapittel 7

Konklusjoner

7.1 Bemerkninger

Vekst i levealder har forårsaket en kraftig økning på engangspremier. Relativ feil i pensjonsporteføljene reduseres med tiden. Det er primært forårsaket av reduksjon i modellfeil siden tilfeldig feil viste seg å være konstant. I 2009 utgjorde gjennomsnittlig modellfeil kun 0.5 % av korrekt engangspremie i landene Norge, Sverige, Japan og Italia. For Russland var feilen fire ganger så stor. Resultatene er i hovedsak beregnet på grunnlag av en statisk Gompertz-Makeham modell. Dette er hverken realistisk eller økonomisk gunstig. Livselskaper står ovenfor store utfordringer i fremtiden. Det må uten tvil tas hensyn til lengre levetid og dynamiske dødelighetsmodeller.

7.2 Veien videre

Planen var å anvende teorien fra seksjon 3.1 og 3.2 på henholdsvis Lee-Carter og Beta-modellen. I Lee-Carter oppstod det problemer siden den stokastiske tidsrekken k_k ikke er en lineær tilstandsmodell[8]. Det var synd, og førte til begrenset dynamisk dødelighetsanalyse. Derimot, i beta-modellen oppstod det problemer i forbindelse med avrundingsfeil ved optimering av likelihood funksjonen når $N \rightarrow \infty$. Det var ikke nok tid til å løse dette problemet. Et naturlig skritt videre er å undersøke avrundingsfeilen nærmere. Mest sannsynlig er programspråket R for svakt slik at andre programmer vil være nødvendig. For eksempel Fortran eller C++.

Del III

Vedlegg

Kapittel 8

Vedlegg

8.1 Vedlegg A - Formler

Dødelighetsnotasjon

Her følger en oversikt over standard dødelighetsnotasjon

- q_x : Sannsynligheten å dø i aldersintervallet $(x, x+1)$
- p_x : Sannsynligheten å overleve i aldersintervallet $(x, x+1)$
- n_x : Antall gjenlevende i alder x , baseres på en index. Matematisk, $n_{x+1} = n_x \cdot p_x$ eller $p_x = \frac{n_{x+1}}{n_x}$
- d_x : Antallet som dør i en bestemt alder x . Matematisk $d_x = n_x - n_{x+1}$
- μ_x : Dødelighetsfrekvens i alder x . Det er en frekvens som viser antall døende i en veldig kort periode. Brukes for eksempel i Gompertz-Makeham.
- ${}_k p_x$: Sannsynligheten for å overleve k år frem i tid ved alder x . Matematisk ${}_k p_x = \frac{l_{x+k}}{l_x}$.

Maximum Likelihood-estimator

La X_1, X_2, \dots, X_n ha identisk sannsynlighetstetthet

$$f(x_1, x_2, \dots, x_n : \theta_1, \theta_2, \dots, \theta_n)$$

hvor parametrene $\theta_1, \theta_2, \dots, \theta_m$ har ukjente verdier. Når x_1, x_2, \dots, x_n er observerte verdier som en funksjon av $\theta_1, \theta_2, \dots, \theta_m$ er (8.1) en **likelihood funksjon**. Vi kaller $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m$ *maximum likelihood estimatorene* som maksimerer alle θ_i slik at [5]

$$f(x_1, x_2, \dots, x_n : \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m) \geq f(x_1, x_2, \dots, x_n : \theta_1, \theta_2, \dots, \theta_n)$$

Gamma-funksjon

En Gamma-funksjonen er definert[5]

$$\Gamma(\alpha) = \int_0^{\infty} e^{-u} u^{\alpha-1} du$$

og har egenskaper

$$\Gamma(\alpha + 1) = \Gamma(\alpha) \cdot \alpha, \quad \Gamma(n) = (n - 1)! \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

hvor n må være et heltall

Beta-funksjon

En Beta-funksjonen er definert[5]

$$\beta(x, y) = \int_0^1 u^{x-1} \cdot (1 - u)^{y-1} du$$

og har egenskap

$$\beta(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

Sammensatt sannsynlighetsfordeling

En sammensatt sannsynlighets-fordeling resulterer å anta at de tilfeldige variablene er fordelt ved en parametrisk fordeling med en ukjent parameter θ som tilhører en annen sannsynlighetsfordeling. Anta at det finnes to slike fordelinger, F og G, vil en kontinuerlig sammensatt sannsynlighetsfordeling være gitt ved[5]

$$p_H(x/\theta) = \int_{\theta} p_F(x/\theta) \cdot p_G(\theta/\alpha)$$

8.2 Vedlegg B - Programkode

8.2.1 Kapittel 2

```
1 #Dødelighetsdata
2 Z = matrix(read.table("C:/Users/Lars/Desktop/Russiskemenn/nor2table60-09.txt"),
  ,byrow=T,ncol=10)
3 K = Z[,6][[1]]
4 U = matrix(read.table("C:/Users/Lars/Desktop/Russiskemenn/1959-2010.txt"),
  ,byrow=T,ncol=10)[6][[1]]
5 ll = 111
6 V = matrix(U,byrow=T,ncol=11)
7 Z = matrix(0,5,11)
8
9 for (i in 1:5){
10 Z[i,]=apply(V[(10*(i-1)+2):(10*i+2)],2,sum)
11 }
12 year = 50;l2 = 110;gx = matrix(0,12,year)
13 for (k in 1:year){
14 G = K[(1+111*(k-1)):(111*k)][1:111]
15 le = length(G)
16 gx[,k] = (1-(G[-1]/G[1:(le-1)]))
17 }
18 #Plot Norge
19 expected = 0.5+apply(apply(1-gx,2,cumprod),2,sum)
20 plot(ts(expected,start=1960),lwd=3,ylab="Alder",xlab="Periode"
21 ,main="Forventet levetid ved fødsel for norske menn")
22
23 #Plot Russland
24 library(ggplot2)#R-package
25 Land = c(rep("Russland",51),rep("Norge",51))
26 Relativframe = c(ZZ[1:51],QQ)
27 frame = data.frame(Relativframe,Land)
28 p=qplot(rep(1959:2009,2),Relativframe,
29 data = frame,geom = c("line"),colour=Land,
30 xlab="Alder",ylab="Antall Gjenlevende"
31 ,main="Tiårs dødelighetstabeller, russiske menn")
32 p + opts(axis.title.y=theme_text(angle=90)
33 ,panel.background = theme_rect(fill='white'
34 , colour="black"))
35
36 Periode = c(rep("1960-1969",31),rep("1970-1979",31),rep("1980-1989",31),rep("
  1990-1999",31),rep("2000-2009",31))
37 Relativframe = c(R[1,][40:70],R[2,][40:70],R[3,][40:70],R[4,][40:70],R
  [5,][40:70])
38 frame = data.frame(Relativframe,Periode)
39 p=qplot(rep(40:70,5),Relativframe, data = frame,geom = c("line"),colour=
  Periode,
40 xlab="Alder",ylab="Antall Gjenlevende",main="Tiårs dødelighetstabeller,
  russiske menn")
41 p + opts(axis.title.y=theme_text(angle=90),panel.background = theme_rect(fill=
  'white', colour="black"))+ labs(colour = "Periode")
```

Kapittel2.R

8.2.2 Kapittel 4

```
1 library(demography)#R-pakke, HMD
2
```

8.2. VEDLEGG B - PROGRAMKODE

```
3 #Bruker R-pakken til å hente data fra HMD
4 NorMx <- hmd.mx("NOR", username="email", password="xxx","Norway")
5 RusMx <- hmd.mx("RUS", username="email", password="xxx","Russia")
6 JapMx <- hmd.mx("JPN", username="email", password="xxx","Japan")
7 UsaMx <- hmd.mx("USA", username="email", password="xxx","Usa")
8 SweMx <- hmd.mx("SWE", username="email", password="xxx","Sweden")
9 ItaMx <- hmd.mx("ITA", username="email", password="xxx","Italy")
10
11 l1 = 1:103
12 year = 51
13 Y1=NorMx$pop$male[l1,114:164]
14 rate1=NorMx$rate$male[l1,114:164]
15 Z1=Y1*rate1
16
17 Y2=RusMx$pop$male[l1,1:51]
18 rate2=RusMx$rate$male[l1,1:51]
19 Z2=Y2*rate2
20
21 Y3=JapMx$pop$male[l1,13:63]
22 rate3=JapMx$rate$male[l1,13:63]
23 Z3=Y3*rate3
24
25 Y4=UsaMx$pop$male[l1,28:78]
26 rate4=UsaMx$rate$male[l1,28:78]
27 Z4=Y4*rate4
28
29 Y5=SweMx$pop$male[l1,211:261]
30 rate5=SweMx$rate$male[l1,211:261]
31 Z5=Y5*rate5
32
33 Y6=ItaMx$pop$male[l1,88:138]
34 rate6=ItaMx$rate$male[l1,88:138]
35 Z6=Y6*rate6
36
37 #Forventet Levalder
38 Forv.Nor = apply(apply(1-rate1,2,cumprod),2,sum)
39 Forv.Rus = apply(apply(1-rate2,2,cumprod),2,sum)
40 Forv.Usa = apply(apply(1-rate4,2,cumprod),2,sum)
41 Forv.Jap = apply(apply(1-rate3,2,cumprod),2,sum)
42 Forv.Ita = apply(apply(1-rate6,2,cumprod),2,sum)
43 Forv.Swe = apply(apply(1-rate5,2,cumprod),2,sum)
44
45 #Plot
46 library(ggplot2)
47 gammaframe = c(Forv.Nor,Forv.Rus,Forv.Usa,Forv.Jap,Forv.Ita,Forv.Swe)
48 Land = c(rep("1",51),rep("2",51),
49 rep("3",51),rep("4",51),rep("5",51),rep("6",51))
50 frame = data.frame(gammaframe,Land)
51 p = qplot(rep(1959:2009,6),gammaframe, data = frame, geom=c("line"),colour=
52 Land,xlab="Periode",
53 ylab="Alder",main="Forventet levalder for nyfødte")
54 p + scale_colour_manual(name = "",
55 labels = c("Norge", "Russland", "Usa", "Japan","Italia",
56 "Sverige"),
57 values = c(1:6)) + guides(col = guide_legend(reverse = TRUE))+ opts(axis.title
58 .y=theme_text(angle=90)
59 ,panel.background = theme_rect(fill='white', colour="black"),legend.text =
60 theme_text(size=12))+ labs(colour = "")
```


8.2.3 Kapittel 5

5.1 + 5.2 - 2009

```
1 #Data norske og russiske menn fra 2009
2 Z = matrix(read.table("C:/Users/Lars/Desktop/Russiskemenn/rustable09.txt"),
  byrow=T,ncol=10)
3 Q = matrix(read.table("C:/Users/Lars/Desktop/Russiskemenn/nortable09.txt"),
  byrow=T,ncol=10)
4 G = Z[,6][[1]]#Russisk
5 T = Q[,6][[1]]#Norsk
6 xe = length(G)
7 gl1 = 1-(G[-1]/G[1:(xe-1)]);gl1[110] = 1#Russisk
8 gl2 = 1-(T[-1]/T[1:(xe-1)])#Norsk
9 #Tabell 5.1
10 round(gl1[c(1,10,30,50,75)]*100,3)
11 round(gl2[c(1,10,30,50,75)]*100,3)
12 l1 = 1;l2 = length(gl1);l1=l1:l2
13
14 #MLE optimering for Gompertz-Makeham
15 minusgompakll1=function(s,l1,n,z)
16 {
17 t=exp(s)
18 p=exp(-t[1]-t[2]*exp(t[3]*l1))
19 -sum((n-z)*log(p)+z*log(1-p))
20 }
21 gompakfit1=function(l1,n,z)
22 {
23 s=-c(8,9,2)
24 o=optim(s,minusgompakll1,"BFGS",l1=l1,n=n,z=z)
25 theta=exp(o$par)
26 list(theta=theta)
27 }
28
29 #Funksjon for Engangspremie
30 onetimeprem = function(q,s,l_r,d)
31 {
32 l_e = length(q)
33 kp = lifetab(q,l_e)$kp
34 l = matrix(0,l_e+1,l_e+1)
35 l[row(l)+col(l)>l_r+1]=1
36 ll=0:l_e
37 M = s*(d**ll)*kp*l
38 pi = apply(M,2,sum)
39 list(pi=pi)
40 }
41 lifetab = function(q,K)
42 {
43 l_e = length(q)
44 p = c(1-q,rep(0,l_e+1))
45 kp = matrix(1,K,l_e+1)
46 for (l in 0:l_e){
47 kp[,l] = cumprod(p[1:K+(l-1)])
48 }
49 kp = rbind(rep(1,l_e+1),kp)
50 list(kp=kp)
51 }
52
53 #Figur 5.1 a)
54 library(ggplot2)
55 xx = 1:75;Norge = gl2[xx];Russland = gl1[xx]
```

8.2. VEDLEGG B - PROGRAMKODE

```

56 Trueframe = c(Norge,Russland);Land = c(rep("1",length(xx)),rep("2",length(xx))
57 )
58 frame2 = data.frame(Trueframe,Land)
59 p=qplot(rep(xx,2),Trueframe, data = frame2,geom = c("auto"),colour=Land,
60 xlab="Alder",ylab="Dødelighet",main=expression(paste("g"[x]("Norge")," vs","
61 g"[x]("Russland"))))
62 p + scale_colour_manual(
63 labels = c(expression("g"[x]: "Norge"),expression(
64 "g"[x]: "Russland")),values = colors()[c(134,24)]+
65 opts(axis.title.y=theme_text(angle=90)
66 ,legend.background = theme_rect(fill = colors()[1], colour = NA),panel.
67 background = theme_rect(fill='white', colour='black'))
68 #Figur 5.1 b)
69 frame = data.frame(Russland,Norge)
70 p = ggplot(frame,aes(x=Norge,y=Russland),main="QQ-plot") + geom_point()
71 p + coord_cartesian(
72 ylim = c(-0.005, 0.11)) + opts(axis.title.y=theme_text(angle=90),
73 title = "QQ-plot",panel.background = theme_rect(fill='white',colour='black'))
74 + labs(y=expression("g"[x]: "Russland"),x=expression("g"[x]: "Norge"))
75 #Aldersprofil
76 gamma = 0.10;lbar = 40;l = 0:109;c = 1/sum(exp(-(gamma*abs(l-lbar))))
77 sumw1 = c*sum(exp(-(gamma*abs(l-lbar))));w1 = c*exp(-(gamma*abs(l-lbar)));n1 =
78 w1
79 y11 = n1*g11;y12 = n1*g12
80 #Kullback-Leibler minimering
81 thetahatrus = gompakfit1(l1=l1,n=n1,z=y11)$theta;thetahatnor = gompakfit1(l1
82 =l1,n=n1,z=y12)$theta
83 #Kullback-Leibler dødelighet
84 qlrus = 1-exp(-thetahatrus[1]-thetahatrus[2]*exp(thetahatrus[3]*l1))
85 qlnor = 1-exp(-thetahatnor[1]-thetahatnor[2]*exp(thetahatnor[3]*l1))
86 #Tabell 5.2
87 round(qlrus[c(1,10,30,50,75)]*100,3);round(qlnor[c(1,10,30,50,75)]*100,3)
88 setwd("C:/Users/Lars/Desktop/ilc");attach("ilc.rdata");attach("cmi.rdata");
89 args(lca.rh)#R-Pakke
90 #Figur 5.2
91 par(mfrow=c(1,1));plot(log(g12[xx]),lwd=1,ylim=c(-10,-2.5),ylab="Logaritmisk
92 dødelighet",xlab="Alder",main=expression(paste(q[theta[0]][x]," vs"," g"
93 [x])),col="blue")
94 lines(ts(log(g11[xx]),start=1),"p",col="blue",lwd=1);lines(log(qlnor[xx]),lwd
95 =3,col="black")
96 lines(log(qlrus[xx]),lwd=3,col="black");legend(coord("UL"), legend=c(
97 expression("g"[x]),expression("q"[theta[0]][x])),
98 lty = c(1,NA),pch=c(NA,1),col = c("1","blue"),lwd=c(3,1));text(30, -4.5, "
99 Russland");text(45, -7, "Norge")
100 ###Bootstrap###
101 ###BOOTSTRAP
102 NN = c(10000,100000,500000,1000000,10000000,10000000000000)
103 nn = length(NN);mb = 500
104 thetarus = array(0,dim=c(3,mb,nn));thetator = array(0,dim=c(3,mb,nn))
105 qhatrus = array(0,dim=c(12-l1+1,mb,nn));qhatnor = array(0,dim=c(12-l1+1,mb,nn)
106 )
107 for (k in 1:nn){
108 N = NN[k];n1 = round(N*w1)
109 Zrus = matrix(rbinom(mb*(12-l1+1),n1,g11),12-l1+1,mb);Znor = matrix(rbinom(mb*
110 (12-l1+1),n1,g12),12-l1+1,mb)
111 for (i in 1:mb){
112 thetarus[,i,k] = gompakfit1(l1,n1,Zrus[,i])$theta;thetator[,i,k] =
113 gompakfit1(l1,n1,Znor[,i])$theta
114 qhatrus[,i,k] = 1-exp(-thetarus[1,i,k]-thetarus[2,i,k]*exp(thetarus[3,i,k]*l1)
115 )

```

8.2. VEDLEGG B - PROGRAMKODE

```

100 qhatnor[,i,k] = 1-exp(-thetanor[1,i,k]-thetanor[2,i,k]*exp(thetanor[3,i,k]*l1)
    )})
101 #Figur 5.3
102 par(mfrow=c(1,2))
103 l1 = 1:75
104 matplot(l1,cbind(log(qhatrus[l1,,1]),log(qhatnor[l1,,1])), "l",xlab="Alder",
    ylab="Logaritmisk skala",main="N = 10 000")
105 matplot(l1,cbind(log(qhatrus[l1,,4]),log(qhatnor[l1,,3])), "l",xlab="Alder",
    ylab="Logaritmisk skala",main="N = 1 000 000")
106 Mean.Rus = apply(qhatrus[,4],1,mean)
107 Mean.Nor = apply(qhatnor[,4],1,mean)
108 Sd.Rus = apply(qhatrus[,4],1,sd)
109 Sd.Nor = apply(qhatnor[,4],1,sd)
110 #Tabell 5.3
111 round(Mean.Rus[c(1,10,30,50,75)]*100,3)
112 round(Mean.Nor[c(1,10,30,50,75)]*100,3)
113 #Figur 5.4
114 Forhold.Rus = round(Mean.Rus[c(1,10,30,50,75)]/Sd.Rus[c(1,10,30,50,75)],3)
115 Forhold.Nor = round(Mean.Nor[c(1,10,30,50,75)]/Sd.Nor[c(1,10,30,50,75)],3)
116 FR = (Mean.Rus/Sd.Rus)[xx];FN = (Mean.Nor/Sd.Nor)[xx]
117 SR = ((Mean.Rus-qlrus)/qlrus)[xx];SN = ((Mean.Nor-qlnor)/qlnor)[xx]
118 data = c(FR,FN)
119 Land = c(rep("1",length(l1)),rep("2",length(l1)))
120 frame = data.frame(Land,data)
121 p=qplot(rep(l1,2),data, data = frame,geom = c("auto"),colour=Land,
122 xlab="Alder",ylab=expression(frac(E(hat(q)[theta]),sd(hat(q)[theta]))),main="
    Forhold")
123 p + scale_colour_manual(
124     labels = c("Russland","Norge"),
125     values = colors()[c(134,592)]+ opts(axis.title.y=theme_text(angle=0)
126 ,legend.background = theme_rect(fill = colors()[1], colour = NA)
127 ,panel.background = theme_rect(fill='white', colour='black'))
128 data = c(SR,SN)*100
129 Land = c(rep("1",length(l1)),rep("2",length(l1)))
130 frame = data.frame(Land,data)
131 p=qplot(rep(l1,2),data, data = frame,geom = c("auto"),colour=Land,ylim=c
    (-0.5,5),
132 xlab="Alder",ylab="%",main="Skjevhet i prosent")
133 p + scale_colour_manual(
134     labels = c("Russland","Norge"),
135     values = colors()[c(134,592)]+ opts(axis.title.y=theme_text(angle=0)
136 ,legend.background = theme_rect(fill = colors()[1], colour = NA)
137 ,panel.background = theme_rect(fill='white', colour='black'))
138 #Tabell 5.4
139 mean.thetarus.million = apply(thetarus[,4],1,mean);mean.thetanor.million =
    apply(thetanor[,4],1,mean)
140 sd.thetarus.million = apply(thetarus[,4],1,sd);sd.thetanor.million = apply(
    thetanor[,4],1,sd)
141 Forholdrus = mean.thetarus.million/sd.thetarus.million;Forholdnor = mean.
    thetanor.million/sd.thetanor.million
142 Skjevhetrus = round(100*(mean.thetarus.million-thetahatrus)/thetahatrus,2)
143 Skjevhetnor = round(100*(mean.thetanor.million-thetahatnor)/thetahatnor,2)
144 round(Forholdrus,1);round(Forholdnor,1);Skjevhetrus;Skjevhetnor
145 #Avvik i engangspremie
146 lr=65;l_r = 65;s = 150000;r = 0.03; d = 1/(1+r)
147 Pi.nor = onetimeprem(gl2,s,l_r,d)$pi;Pi.rus = onetimeprem(gl1,s,l_r,d)$pi
148 Pi0.nor = onetimeprem(qlnor,s,l_r,d)$pi;Pi0.rus = onetimeprem(qlrus,s,l_r,d)$
    pi
149 sysNorge = abs(Pi.nor-Pi0.nor)[1:65];sysRussland = abs(Pi.rus-Pi0.rus)[1:65]
150 RSNorge = (sysNorge/Pi.nor[1:65])*100;RSRussland = (sysRussland/Pi.rus[1:65])*
    100
151 Pihat.rus = array(0,dim=c(12-l1+2,mb,nn));Pihat.nor = array(0,dim=c(12-l1+2,mb
    ,nn))

```

8.2. VEDLEGG B - PROGRAMKODE

```
152 RKFNorge = array(0,dim=c(65,mb,nn));RKFRussland = array(0,dim=c(65,mb,nn))
153 RKFNorge = matrix(0,65,nn);RKFRussland = matrix(0,65,nn)
154 RTFNorge = matrix(0,65,nn);RTFRussland = matrix(0,65,nn)
155 for (k in 1:nn){
156   for (i in 1:mb){
157     Pihat.nor[,i,k] = onetimeprem(qhatnor[,i,k],s,l_r,d)$pi
158     Pihat.rus[,i,k] = onetimeprem(qhatrus[,i,k],s,l_r,d)$pi
159   }
160   RKFNorge[,k] = sqrt(apply((Pihat.nor[, ,k]-Pi0.nor)**2+(Pi.nor-Pi0.nor)**2,1,
161     mean))[1:65]
161   RKFRussland[,k] = sqrt(apply((Pihat.rus[, ,k]-Pi0.rus)**2+(Pi.rus-Pi0.rus)**
162     2,1,mean))[1:65]
162   RTFNorge[,k] = sqrt(apply((Pihat.nor[, ,k]-Pi0.nor)**2,1,mean)[1:65])/Pi.nor
163     [1:65]
163   RTFRussland[,k] = sqrt(apply((Pihat.rus[, ,k]-Pi0.rus)**2,1,mean)[1:65])/Pi.rus
164     [1:65]
164 }
165 #Figur 5.12
166 par(mfrow=c(1,1))
167 plot(sysNorge,lty=2,ylab="Kroner",xlab="Alder",col="blue",ylim=c(0,30000),main
168   ="Feil i engangspremie, N = 1 000 000")
168 lines(ts(sysRussland,start=1),"p",col="blue");lines(RKFNorge[,4],col="black");
169   lines(RKFRussland[,4],col="black")
169 legend(coord("UL"), legend=c("Systematisk feil","RKF"),lty = c(NA,1),pch=c(1,
170   NA),col = c("blue","1"),lwd=c(1,1))
170 text(40, 12000, "Russland");text(55, 5500, "Norge")
171 #Figur 5.13a)
172 Land = c(rep("Norge",65),rep("Russland",65))
173 Relativframe = c(RSNorge,RSRussland)
174 frame = data.frame(Relativframe,Land)
175 p=qplot(rep(1:65,2),Relativframe, data = frame,geom = c("auto"),colour=Land,
176   xlab="Alder",ylab="%",main="Relativ systematisk feil i prosent")
177 p + scale_colour_manual(
178   labels = c("Norge", "Russland"),
179   values = colors()[c(592,134)]+ opts(axis.title.y=theme_text(angle=0)
180   ,legend.background = theme_rect(fill = colors()[1], colour = NA)
181   ,panel.background = theme_rect(fill='white', colour='black'),legend.text =
182     theme_text(size=14))
182 #Figur 5.13b)
183 Land = c(rep("1",195),rep("2",195))
184 Tilfeldigframe = c(RTFNorge[,c(2,4,5)],RTFRussland[,c(2,4,5)])*100
185 frame = data.frame(Tilfeldigframe)
186 p=qplot(rep(1:65,6),Tilfeldigframe, data = frame,geom = c("point"),colour=Land
187   ,
188   xlab="Alder",ylab="%",main="Relativ tilfeldig feil i prosent")
188 p + scale_colour_manual(
189   labels = c("Norge", "Russland"),
190   values = colors()[c(592,134)]+ opts(axis.title.y=theme_text(angle=0)
191   ,legend.background = theme_rect(fill = colors()[1], colour = NA)
192   ,panel.background = theme_rect(fill='white', colour='black')
193   ,legend.text = theme_text(size=14))+annotate("text"
194   , label = c("N = 100 000","N = 1 000 000", "N = 10 000 000"), x = c(35,35,35)
195   ,y = c(2.8,1.2,0.45), size = 4)+ labs(colour = "")
196 #Tabell 5.12
197 mean(RSNorge);mean(RSRussland);sd(RSNorge);sd(RSRussland)
198 ForNor=(apply(Pihat.nor[, ,4],1,mean)/apply(Pihat.nor[, ,4],1,sd))[1:65]
199 ForRus=(apply(Pihat.rus[, ,4],1,mean)/apply(Pihat.rus[, ,4],1,sd))[1:65]
200 SkjevNor0=(apply(Pihat.nor[, ,4]-Pi0.nor,1,mean)/Pi0.nor)[1:65]
201 SkjevRus0=(apply(Pihat.rus[, ,4]-Pi0.rus,1,mean)/Pi0.rus)[1:65]
202
203 #Figur 5.14a)
204 Land = c(rep("Norge",65),rep("Russland",65))
205 Relativframe = c(ForNor,ForRus)
```

8.2. VEDLEGG B - PROGRAMKODE

```
206 frame = data.frame(Relativframe, Land)
207 p=qplot(rep(1:65,2), Relativframe, data = frame, geom = c("auto"), colour=Land,
208 xlab="Alder", ylab=expression(frac(E(hat(psi)), Sd(hat(psi)))), main="Forhold")
209 p + scale_colour_manual(
210     labels = c("Norge", "Russland"),
211     values = colors()[c(592,134)] + opts(axis.title.y=theme_text(angle=0)
212 , panel.background = theme_rect(fill='white', colour="black")) + annotate("text"
213 , label = c("N = 1 000 000"), x = c(10)
214 , y = c(178.5), size = 5) + labs(colour = "")
215 #Figur 5.14b)
216 Relativframe = c(SkjevNor0, SkjevRus0)*100
217 frame = data.frame(Relativframe, Land)
218 p=qplot(rep(1:65,2), Relativframe, data = frame, geom = c("auto"), colour=Land,
219 xlab="Alder", ylab="%", main="Skjevhet")
220 p + scale_colour_manual(
221     labels = c("Norge", "Russland"),
222     values = colors()[c(592,134)] + opts(axis.title.y=theme_text(angle=0)
223 , panel.background = theme_rect(fill='white', colour="black"), legend.text =
224     theme_text(size=14) + annotate("text"
225 , label = c("N = 1 000 000"), x = c(10)
226 , y = c(0.12), size = 5) + labs(colour = "")
227 SysNorProsent = matrix(0,65,nn); SysRusProsent = matrix(0,65,nn)
228 for (i in 1:nn){
229     SysNorProsent[,i] = sysNorge/RKFNorge[,i]; SysRusProsent[,i] = sysRussland/
230     RKFRussland[,i]
231 }
232 #Figur 5.15a)
233 systematiskframe = c(SysNorProsent)
234 Eksponeringsstørrelse = c(rep("1",65), rep("2",65), rep("3",65), rep("4",65), rep(
235     "5",65), rep("6",65))
236 frame = data.frame(systematiskframe, Eksponeringsstørrelse)
237 p=qplot(rep(1:65,6), systematiskframe, data = frame, geom = c("line"), colour=
238     Eksponeringsstørrelse,
239 xlab="Alder", ylab="%", main="Systematisk feil i prosent, Norge")
240 p + scale_colour_manual(name = "Eksponeringsstørrelse",
241     labels = c("N = 10,000", "N = 100,000", "N = 500,000", "
242     N = 1,000,000", "N = 10,000,000", "N > 50,000,000"),
243     values = c(1:6)) + guides(col = guide_legend(reverse = TRUE)) + opts(axis.
244     title.y=theme_text(angle=0)
245 , legend.background = theme_rect(fill = colors()[1], colour = NA)
246 , panel.background = theme_rect(fill='white', colour='black'))
247 #Figur 5.15b)
248 systematiskframe = c(SysRusProsent)
249 frame = data.frame(systematiskframe, Eksponeringsstørrelse)
250 p=qplot(rep(1:65,6), systematiskframe, data = frame, geom = "line", colour=
251     Eksponeringsstørrelse,
252 xlab="Alder", ylab="%", main="Systematisk feil i prosent, Russland")
253 p + scale_colour_manual(name = "Eksponeringsstørrelse",
254     labels = c("N = 10,000", "N = 100,000", "N = 500,000", "
255     N = 1,000,000", "N = 10,000,000", "N > 50,000,000"),
256     values = c(1:6)) + guides(col = guide_legend(reverse = TRUE)) + opts(axis.title
257     .y=theme_text(angle=0)
258 , legend.background = theme_rect(fill = colors()[1], colour = NA)
259 , panel.background = theme_rect(fill='white', colour='black'))
260 ##Modellfeil i pensjonsportefølje
261 N = 45000; l1 = 40; l2=110; xx=l1:l2; lr=65
262 x1 = 40; x2 = 110; artilpensjon = lr-x1
263 kpxnor = cumprod(1-gl2[x1:x2]); kpxrus = cumprod(1-gl1[x1:x2])
264 kpqnor = cumprod(1-qlnor[x1:x2]); kpqrus = cumprod(1-qlrus[x1:x2])
265 d = (1/(1+0.03))**(25:71)
266 kp = kpxnor[25:71]
267 kq = kpqnor[25:71]
268 kpr = kpxrus[25:71]
```

8.2. VEDLEGG B - PROGRAMKODE

```
260 kqr = kpqrus[25:71]
261 s = 150000
262 pinor = s*sum(d*kp)
263 pi0nor = s*sum(d*kq)
264 pirus = s*sum(d*kpr)
265 pi0rus = s*sum(d*kqr)
266 NN = 10**9
267 VnorTrue = pinor*N/NN;Vprenor1 = VnorTrue*1.03^(1:(lr-x1))
268 VnorKL = pi0nor*N/NN;Vprenor2 = VnorKL*1.03^(1:(lr-x1))
269 VrusTrue = pirus*N/NN;Vprerus1 = VrusTrue*1.03^(1:(lr-x1))
270 VrusKL = pi0rus*N/NN;Vprerus2 = VrusKL*1.03^(1:(lr-x1))
271 enor = N*kpxnor;erus = N*kpxrus;NN = 10^9;s = 150000/NN
272 gjenlevendeNor = enor[-(1:(årtilpensjon-1))]
273 gjenlevendeRus = erus[-(1:(årtilpensjon-1))]
274 V1Nor = rep(Vprenor1[25],length(gjenlevendeNor));VNor = rep(Vprenor2[25],
length(gjenlevendeNor))
275 V1Rus = rep(Vprerus1[25],length(gjenlevendeRus));VRus = rep(Vprerus2[25],
length(gjenlevendeRus))
276 for (i in 1:(length(gjenlevendeNor)-1)){
277   if (V1Nor[i]>0)
278     V1Nor[i+1] = (V1Nor[i] - gjenlevendeNor[i]*s)*(r+1)
279   else
280     V1Nor[i+1] = V1Nor[i] - gjenlevendeNor[i]*s
281   if (VNor[i]>0)
282     VNor[i+1] = (VNor[i] - gjenlevendeNor[i]*s)*(r+1)
283   else
284     VNor[i+1] = VNor[i] - gjenlevendeNor[i]*s
285   if (V1Rus[i]>0)
286     V1Rus[i+1] = (V1Rus[i] - gjenlevendeRus[i]*s)*(r+1)
287   else
288     V1Rus[i+1] = V1Rus[i] - gjenlevendeRus[i]*s
289   if (VRus[i]>0)
290     VRus[i+1] = (VRus[i] - gjenlevendeRus[i]*s)*(r+1)
291   else
292     VRus[i+1] = VRus[i] - gjenlevendeRus[i]*s
293 }
294 VTrue.nor = c(Vprenor1[-25],V1Nor)
295 V0.nor = c(Vprenor2[-25],VNor)
296 VTrue.rus = c(Vprerus1[-25],V1Rus)
297 V0.rus = c(Vprerus2[-25],VRus)
298 Modell = c(rep("1",71),rep("2",71))
299 DynamiskNor = c(VTrue.nor,V0.nor)
300 DynamiskRus = c(VTrue.rus,V0.rus)
301 frameNor = data.frame(DynamiskNor,Modell)
302 frameRus = data.frame(DynamiskRus,Modell)
303 #Figur 5.16
304 p=qplot(rep(2009:2079,2),DynamiskNor, data = frameNor,geom = c("line"),colour=
Modell,
305 xlab="Periode",ylab="Milliarder kroner",main="Akkumulert porteføljeverdi,
norske menn")
306 p + annotate("text", label = c("Utbetalingsperiode","Investeringsperiode"), x
= c(2064,2022),y = c(50,25), size = 5) + scale_colour_manual(
307 labels = c(expression(Psi[Korrekt]),expression(Psi[
gomperz])),
308 values = c(1:2)) + opts(axis.title.y=theme_text(angle=90)
309 ,legend.background = theme_rect(fill = colors()[1], colour = NA)
310 ,panel.background = theme_rect(fill='white', colour='black'),legend.text =
theme_text(size=13))
311 p=qplot(rep(2009:2079,2),DynamiskRus, data = frameRus,geom = c("line"),colour=
Modell,
312 xlab="Periode",ylab="Milliarder kroner",main="Akkumulert porteføljeverdi,
russiske menn")
```

8.2. VEDLEGG B - PROGRAMKODE

```
313 p + annotate("text", label = c("Utbetalingsperiode", "Investeringsperiode"), x
    = c(2064, 2022), y = c(21, 11), size = 5) + scale_colour_manual(
314     labels = c(expression(Psi[Korrekt]), expression(Psi[
        Gomperz])),
315     values = c(1:2)) + opts(axis.title.y=theme_text(angle=90)
316     , legend.background = theme_rect(fill = colors()[1], colour = NA)
317     , panel.background = theme_rect(fill='white', colour='black'), legend.text =
        theme_text(size=13))
318 #Tabell 5.10
319 c(VTrue.nor[c(25, 35, 45, 55, 71)]); c(VTrue.rus[c(25, 35, 45, 55, 71)])
320 c(V0.nor[c(25, 35, 45, 55, 71)]); c(V0.rus[c(25, 35, 45, 55, 71)])
321 gl = gl2*0.85; s=150000; d=1/(1+r)
322 Pi.average.nor = ts(onetimeprem(gl2, s, l_r, d)$pi, start=0); Pi.reduced.nor = ts(
    onetimeprem(gl, s, l_r, d)$pi, start=0)
323 Pi.average.nor[c(30, 40, 50, 65, 80)]; Pi.reduced.nor[c(30, 40, 50, 65, 80)]; Pi.gomperz
    .nor[c(30, 40, 50, 65, 80)]
324 zz = 31:81
325 Engang = c(Pi.average.nor[zz], Pi.reduced.nor[zz])
326 Metode = c(rep("1", 51), rep("2", 51)); frame = data.frame(Engang, Metode)
327 #Figur 5.17a)
328 p=qplot(rep(30:80, 2), Engang, data = frame, geom = c("line"), colour=Metode,
329     xlab="Alder", ylab="Kroner", main="Engangspremie i kroner")
330 p + scale_colour_manual(
331     labels = c(expression(Psi[average]), expression(Psi[
        reduced])),
332     values = c(1:2)) + opts(axis.title.y=theme_text(angle=90)
333     , legend.background = theme_rect(fill = colors()[1], colour = NA)
334     , panel.background = theme_rect(fill='white', colour='black'))
335 ##Ugunstig utvalg##
336 glreduced = gl2*0.85
337 kpx2nor = cumprod(1-glreduced[x1:x2])
338 kpreduced = kpx2nor[25:71]
339 #s = 150000
340 #piReduced = s*sum(d*kpreduced)
341 VReduced = pinor*N/NN; Vpre = VReduced*1.03^(1:(lr-x1))
342 eReduced = N*kpx2nor; NN = 10^9; s = 150000/NN
343 gjenlevendeReduced = eReduced[-(1:(artilpensjon-1))]
344 V1 = rep(Vpre[25], length(gjenlevendeNor))
345 for (i in 1:(length(gjenlevendeReduced)-1)){
346     if (V1[i]>0)
347         V1[i+1] = (V1[i] - gjenlevendeReduced[i]*s)*(r+1)
348     else
349         V1[i+1] = V1[i] - gjenlevendeReduced[i]*s
350 }
351 V.All.Reduced = c(Vpre[-25], V1)
352 c(V.All.Reduced[c(25, 35, 45, 55, 71)])
353 Dynamisk = c(VTrue.nor, V.All.Reduced)
354 Modell = c(rep("1", 71), rep("2", 71))
355 frame = data.frame(Dynamisk, Modell)
356 #Figur 5.17b)
357 p=qplot(rep(2009:2079, 2), Dynamisk, data = frame, geom = c("line"), colour=Modell
    ,
358     xlab="Periode", ylab="Milliarder kroner", main="Akkumulert porteføljeverdi")
359 p + annotate("text", label = c("Utbetalingsperiode", "Investeringsperiode"), x
    = c(2060, 2022), y = c(55, 25), size = 5) + scale_colour_manual(
360     labels = c(expression(Psi[average]), expression(Psi[
        reduced])),
361     values = c(1:2)) + opts(axis.title.y=theme_text(angle=90)
362     , legend.background = theme_rect(fill = colors()[1], colour = NA)
363     , panel.background = theme_rect(fill='white', colour='black'), legend.text =
        theme_text(size=13))
364 #Tabell 5.15
365 c(V.All.average[c(25, 35, 45, 55, 71)]); c(V.All.reduced[c(25, 35, 45, 55, 71)])
```

8.2. VEDLEGG B - PROGRAMKODE

```
366 ##Asymptotisk
367 l1 = 1;l2 = length(g11);l1 = 1:l10;y11 = n1*g11;y12 = n1*g12
368 thetahatrus = gompmafit1(l1=l1,n=n1,z=y11)$theta;thetahatnor = gompmafit1(l1
    =l1,n=n1,z=y12)$theta
369 asymptotic = function(l1,N,g1,q1,n1,theta0,theta1,theta2){
370   y1 = n1*g1
371   f1 = expression(y1*log(1-exp(-theta0-theta1*exp(theta2*l1))))+
372   (n1-y1)*log(exp(-theta0-theta1*exp(theta2*l1))))
373   dfl11 = D(f1,"theta0");d2f11 = D(dfl11,"theta0");l11=sum(eval(d2f11))
374   dfl12 = D(f1,"theta1");d2f12 = D(dfl12,"theta1");l12=sum(eval(d2f12))
375   dfl13 = D(f1,"theta2");d2f13 = D(dfl13,"theta2");l13=sum(eval(d2f13))
376   dfl12 = D(f1,"theta0");d2f12 = D(dfl12,"theta1");l12=sum(eval(d2f12))
377   dfl13 = D(f1,"theta0");d2f13 = D(dfl13,"theta2");l13=sum(eval(d2f13))
378   dfl23 = D(f1,"theta1");d2f23 = D(dfl23,"theta2")
379   l23=sum(eval(d2f123));v11 = (n1*g1*(1-g1) + n1**2*(g1-q1)**2)/q1**2
380   I11 = sum(v11);I22 = sum(v11*exp(2*theta2*l1));I33 = sum(v11*exp(2*theta2*l1)*
    theta1^2*l1^2)
381   I12 = sum(v11*exp(theta2*l1));I13 = sum(v11*exp(theta2*l1)*theta1*l1)
382   I23 = sum(v11*exp(2*theta2*l1)*theta1*l1)
383   l1 = c(l11,l12,l13);l2 = c(l12,l22,l23);l3 = c(l13,l23,l33)
384   v1 = c(I11,I12,I13);v2 = c(I12,I22,I23);v3 = c(I13,I23,I33)
385   lamda1 = matrix(c(l1,l2,l3),nrow=3,ncol=3);v = matrix(c(v1,v2,v3),nrow=3,ncol
    =3)
386   inv.lamda = solve(lamda1);covariance = (inv.lamda %*% v %*% inv.lamda);sd.
    theta = sqrt(diag(covariance))
387   list(sd=round(sd.theta,8),cov=covariance)
388 }
389 NN = c(5000,50000,1000000)
390 asymptoticNor = matrix(0,3,3);asymptoticRus = matrix(0,3,3)
391 ##Norge##
392 for (k in 1:3){
393   N = NN[k]
394   n1 = N*w1
395   theta0 = thetahatnor[1]
396   theta1 = thetahatnor[2]
397   theta2 = thetahatnor[3]
398   asymptoticNor[k,] = asymptotic(l1=l1,N=N,g1=g12,q1=qlnor,n1=n1,theta0=theta0,
    theta1=theta1,theta2=theta2)$sd
399   covnor = asymptotic(l1=l1,N=N,g1=g12,q1=qlnor,n1=n1,theta0=theta0,theta1=
    theta1,theta2=theta2)$cov
400 ##Russland##
401   theta0 = thetahatrus[1]
402   theta1 = thetahatrus[2]
403   theta2 = thetahatrus[3]
404   asymptoticRus[k,] = asymptotic(l1=l1,N=N,g1=g11,q1=qlrus,n1=n1,theta0=theta0,
    theta1=theta1,theta2=theta2)$sd
405   covrus = asymptotic(l1=l1,N=N,g1=g12,q1=qlrus,n1=n1,theta0=theta0,theta1=
    theta1,theta2=theta2)$cov
406 }
407 AsympSdN = t(asymptoticNor)
408 AsympSdR = t(asymptoticRus)
```

OneYear.R

5.1 + 5.2 - 1960-2008

```
1 #Data fra 1960 til 2008,norske og russiske menn, HMD
2 Z = matrix(read.table("C:/Users/Lars/Desktop/Russiskemenn/nor2table60-09.txt"),
    ,byrow=T,ncol=10)
3 Q = matrix(read.table("C:/Users/Lars/Desktop/Russiskemenn/rustable60-09.txt"),
    ,byrow=T,ncol=10)
```


8.2. VEDLEGG B - PROGRAMKODE

```
4 K = Z[,6][[1]];T = Q[,6][[1]]
5
6 minusgompma11=function(s,ll,n,z)
7 {
8 #Returns minus the log-likelihood function under the Gomerz-Makheam model
9 #s=log(theta) where theta is the vector c(theta_1,theta_2,theta_3)
10 t=exp(s)
11 p=exp(-t[1]-t[2]*(exp(t[3]*ll)))
12 -sum((n-z)*log(p)+z*log(1-p))
13 }
14 gompmafit1=function(ll,n,z)
15 {# Likelihood estimates of the Gomerz-Makeham parameters
16 s=-c(8,9,2)
17 o=optim(s,minusgompma11,"BFGS",ll=ll,n=n,z=z)
18 theta=exp(o$par)
19 list(theta=theta)
20 }
21 library(demography)#R-pakke
22 setwd("C:/Users/Lars/Desktop/ilc");attach("ilc.rdata");attach("cmi.rdata");
   args(lca.rh)
23 #Populasjon fra 1960 til 2008, norske og russiske menn, HMD
24 NorMx <- hmd.mx("NOR", username="lars.oesthassel@gmail.com", password="XXX",
   Norway")
25 RusMx <- hmd.mx("RUS", username="lars.oesthassel@gmail.com", password="XXX",
   Russland")
26 Norpop = NorMx$pop$male[1:105,115:163];Ruspop = RusMx$pop$male[1:105,2:50]
27 ages = 105;year = 49;ll = 1:ages;ll1=1;ll2=ages
28 mxNor = matrix(0,ages,year);mxRus = matrix(0,ages,year);mxNor0 = matrix(0,ages
   ,year)
29 mxRus0 = matrix(0,ages,year);mxThetaN = matrix(0,3,year);mxThetaR = matrix
   (0,3,year)
30 for (k in 1:year){
31 nl = Norpop[,k]/sum(Norpop[,k]);G = K[(1+111*(k-1)):(111*k)][1:106]
32 le = length(G);mxNor[,k] = (1-(G[-1]/G[1:(le-1)]));gl = mxNor[,k]
33 yl = nl*gl;mxThetaN[,k] = gompmafit1(ll=ll,n=nl,z=yl)$theta;nl = Ruspop[,k]/
   sum(Ruspop[,k])
34 G = T[(1+111*(k-1)):(111*k)][1:106];le = length(G);mxRus[,k] = (1-(G[-1]/G
   [1:(le-1)]));gl = mxRus[,k]
35 yl = nl*gl;mxThetaR[,k] = gompmafit1(ll=ll,n=nl,z=yl)$theta
36 mxNor0[,k] = (1-exp(-mxThetaN[1,k]-mxThetaN[2,k]*exp(mxThetaN[3,k]*ll)))
37 mxRus0[,k] = (1-exp(-mxThetaR[1,k]-mxThetaR[2,k]*exp(mxThetaR[3,k]*ll)))
38 }
39 ExpectedNor0 = ts(0.5+apply(apply(1-mxNor0,2,cumprod),2,sum),start=1960)
40 ExpectedNor = ts(0.5+apply(apply(1-mxNor,2,cumprod),2,sum),start=1960)
41 ExpectedRus0 = ts(0.5+apply(apply(1-mxRus0,2,cumprod),2,sum),start=1960)
42 ExpectedRus = ts(0.5+apply(apply(1-mxRus,2,cumprod),2,sum),start=1960)
43 #Figur 5.5
44 par(mfrow=c(1,1))
45 plot(ExpectedNor,lwd=3,ylim=c(56,79),ylab="Forventet levealder",xlab="Periode"
   ,main="Forventet levealder for nyfødte, Norge og Russland")
46 lines(ExpectedNor0,"p",col="blue");lines(ExpectedRus,lwd=3,col="black");lines(
   ExpectedRus0,"p",col="blue")
47 legend(coord("UL"), legend=c(expression("e"[0][k]),expression("e"[0][k][theta
   [0]])),lty = c(1,NA),pch=c(NA,1),col = c("1","blue"),lwd=c(1,1))
48 text(1980, 65, "Russland");text(1980, 76, "Norge")
49 #Tabell 5.5
50 ExpectedNor[c(1,21,41)];ExpectedNor0[c(1,21,41)];ExpectedRus[c(1,21,41)];
   ExpectedRus0[c(1,21,41)]
51 ExpectedNor[c(10,30,49)];ExpectedNor0[c(10,30,49)];ExpectedRus[c(10,30,49)];
   ExpectedRus0[c(10,30,49)]
52 #Bootstrap
53 mb = 2000
54 N = 1000000
```

8.2. VEDLEGG B - PROGRAMKODE

```
55 mxBootMeanNor = matrix(0,ages,year)
56 mxBootMeanRus = matrix(0,ages,year)
57 mxNorUpBound = matrix(0,ages,year)
58 mxNorLowBound = matrix(0,ages,year)
59 mxRusUpBound = matrix(0,ages,year)
60 mxRusLowBound = matrix(0,ages,year)
61 meanThetaN = matrix(0,year,3);meanThetaR = matrix(0,year,3);sdThetaN = matrix
  (0,year,3)
62 sdThetaR = matrix(0,year,3);ForholdGMnor = matrix(0,year,3);ForholdGMrus =
  matrix(0,year,3)
63 for (l in 1:year){
64   w11 = Norpop[,l]/sum(Norpop[,l])
65   n11 = round(N*w11)
66   Z = matrix(rbinom(mb*(l2-l1+1),n11,mxNor[,l]),l2-l1+1,mb)
67   w1 = Ruspop[,l]/sum(Ruspop[,l])
68   n12 = round(N*w1)
69   Q = matrix(rbinom(mb*(l2-l1+1),n12,mxRus[,l]),l2-l1+1,mb)
70   mxThetaBootNor = matrix(0,ages,mb)
71   mxThetaBootRus = matrix(0,ages,mb)
72   thetaN = matrix(0,mb,3)
73   thetaR = matrix(0,mb,3)
74   for (i in 1:mb){
75     thetaN[i,] = gompmakfit1(l1:l2,n11,Z[,i])$theta
76     thetaR[i,] = gompmakfit1(l1:l2,n12,Q[,i])$theta
77     mxThetaBootNor[,i] = 1-exp(-thetaN[i,1]-thetaN[i,2]*exp(thetaN[i,3]*l1))
78     mxThetaBootRus[,i] = 1-exp(-thetaR[i,1]-thetaR[i,2]*exp(thetaR[i,3]*l1))
79   }
80   meanThetaN[l,] = apply(thetaN,2,mean);meanThetaR[l,] = apply(thetaR,2,mean)
81   sdThetaN[l,] = apply(thetaN,2,sd);sdThetaR[l,] = apply(thetaR,2,sd)
82   mxBootMeanNor[,l] = apply(mxThetaBootNor,1,mean)
83   ForholdGMnor[l,] = mxBootMeanNor[c(20,40,60),l]/apply(mxThetaBootNor,1,sd)[c
    (20,40,60)]
84   mxNorLowBound[,l] = apply(mxThetaBootNor,1,sort)[round(mb*0.025),];
    mxNorUpBound[,l] = apply(mxThetaBootNor,1,sort)[round(mb*0.975),]
85   mxBootMeanRus[,l] = apply(mxThetaBootRus,1,mean)
86   ForholdGMrus[l,] = mxBootMeanRus[c(20,40,60),l]/apply(mxThetaBootRus,1,sd)[c
    (20,40,60)]
87   mxRusLowBound[,l] = apply(mxThetaBootRus,1,sort)[round(mb*0.025),];
    mxRusUpBound[,l] = apply(mxThetaBootRus,1,sort)[round(mb*0.975),]
88 }
89 Elownor = ts(0.5+apply(apply(1-mxNorLowBound,2,cumprod),2,sum),start=1960)
90 Eupnor = ts(0.5+apply(apply(1-mxNorUpBound,2,cumprod),2,sum),start=1960)
91 Elowrus = ts(0.5+apply(apply(1-mxRusLowBound,2,cumprod),2,sum),start=1960)
92 Euprus = ts(0.5+apply(apply(1-mxRusUpBound,2,cumprod),2,sum),start=1960)
93 ExpectedB00TN = ts(0.5+apply(apply(1-mxBootMeanNor,2,cumprod),2,sum),start
  =1960)
94 ExpectedB00TR = ts(0.5+apply(apply(1-mxBootMeanRus,2,cumprod),2,sum),start
  =1960)
95 #Figur 5.6
96 par(mfrow=c(1,2))
97 fcast = cbind(c(1960:2008),Elownor,Eupnor)
98 plot(ExpectedB00TN,main="Norge",xlab="Periode",ylab="Forventet levealder",lwd
  =2,col="red",ylim=c(70.5,78.5))
99 legend(coord("UL"), legend=c(expression("e"[0][k][hat(theta)]),expression("e"
  [0][k][theta[0]])), lty = c(1,NA),pch=c(NA,1),col = c("red", "blue"),lwd=c
  (2,2))
100 filled.gap(fcast,edges=F,app=T);lines(ExpectedB00TN,lwd=2,col="red",lty=1);
  lines(ExpectedNor0,"p",lwd=2,col="blue")
101 text(1980, 74.5, "95 % Konfidensintervall",cex=0.7)
102 fcast = cbind(c(1960:2008),Elowrus,Euprus)
103 plot(ExpectedB00TR,main="Russland",xlab="Periode",ylab="Forventet levealder",
  lwd=2,col="red",ylim=c(57.5,66))
```

8.2. VEDLEGG B - PROGRAMKODE

```

104 legend(coord("UR"), legend=c(expression("e"[0][k][hat(theta)]), expression("e"
    [0][k][theta[0]])), lty = c(1,NA),pch=c(NA,1),col = c("red", "blue"),lwd=c
    (2,2))
105 filled.gap(fcast,edges=F,app=T);lines(ExpectedB00TR,lwd=2,col="red",lty=1);
    lines(ExpectedRus0,"p",lwd=2,col="blue")
106 text(1970, 61,"95 % Konfidensintervall",cex=0.7)
107 #Figur 5.7 + tabell 5.6
108 par(mfrow=c(1,3))
109 plot(ts(meanThetaN[,1],start=1960),main="Parameter a",ylim=c(0.00006,0.0036),
    xlab="Periode",ylab="a")
110 lines(ts(mxThetaN[,1],start=1960),"p",col="blue");lines(ts(meanThetaR[,1],
    start=1960));lines(ts(mxThetaR[,1],start=1960),"p",col="blue")
111 legend(coord("UR"), legend=c(expression(E(hat(a)),a[0])),lty = c(1,NA),pch=c(
    NA,1),col = c("1","blue"),lwd=c(1,1),ncol=1)
112 text(1970, 0.0023, "Russland");text(1970, 0.00085, "Norge")
113 meanThetaN[,1][c(20,40)];mxThetaN[,1][c(20,40)];meanThetaR[,1][c(20,40)];
    mxThetaR[,1][c(20,40)]
114 plot(ts(round(meanThetaR[,2],6),start=1960),ylim=c(0,0.001),main="Parameter b"
    ,xlab="Periode",ylab="b")
115 lines(ts(mxThetaR[,2],start=1960),"p",col="blue");lines(ts(meanThetaN[,2],
    start=1960));lines(ts(mxThetaN[,2],start=1960),"p",col="blue")
116 legend(coord("UL"), legend=c(expression(E(hat(b)),b[0])),lty = c(1,NA),pch=c(
    NA,1),col = c("1","blue"),lwd=c(1,1),ncol=1)
117 text(1980, 0.00035, "Russland");text(2000, 0.00006, "Norge")
118 round(meanThetaN[,2][c(20,40)],7);round(mxThetaN[,2][c(20,40)],7);meanThetaR
    [,2][c(20,40)];mxThetaR[,2][c(20,40)]
119 plot(ts(round(meanThetaR[,3],6),start=1960),ylim=c(0.060,0.12),main="Parameter
    c",xlab="Periode",ylab="c")
120 lines(ts(mxThetaR[,3],start=1960),"p",col="blue");lines(ts(meanThetaN[,3],
    start=1960));lines(ts(mxThetaN[,3],start=1960),"p",col="blue")
121 legend(coord("UL"), legend=c(expression(E(hat(c)),c[0])),lty = c(1,NA),pch=c(
    NA,1),col = c("1","blue"),lwd=c(1,1),ncol=1)
122 text(1980, 0.070, "Russland");text(1980, 0.105, "Norge")
123 round(meanThetaN[,3][c(20,40)],7);round(mxThetaN[,3][c(20,40)],7);meanThetaR
    [,3][c(20,40)];mxThetaR[,3][c(20,40)]
124 #Figur 5.8
125 forholdRus = meanThetaR/sdThetaR;forholdNor = meanThetaN/sdThetaN;Forhold = c(
    forholdRus[,1],forholdNor[,1])
126 Land = c(rep("1",49),rep("2",49));frame = data.frame(Forhold,Land);library(
    ggplot2)
127 p=qplot(rep(1960:2008,2),Forhold, data = frame,geom = c("point"),colour=Land
128 ,xlab="Periode")
129 #,ylab=expression(frac(E(hat(a)),Sd(hat(a))))
130 p + scale_colour_manual(
131     labels = c("Russland","Norge"),
132     values = colors()[c(592,134)]+ opts(axis.title.y=theme_text(angle=90)
133     ,legend.background = theme_rect(fill = colors()[1], colour = NA)
134     ,panel.background = theme_rect(fill='white', colour='black'))+ geom_smooth(
        method="loess",lwd=1)
135 Forhold = c(forholdRus[,2],forholdNor[,2]);frame = data.frame(Forhold,Land)
136 p=qplot(rep(1960:2008,2),Forhold, data = frame,geom = c("point"),colour=Land,
137 xlab="Periode")
138 #,ylab=expression(frac(E(hat(b)),Sd(hat(b))))
139 p + scale_colour_manual(
140     labels = c("Russland","Norge"),
141     values = colors()[c(592,134)]+ opts(axis.title.y=theme_text(angle=90)
142     ,legend.background = theme_rect(fill = colors()[1], colour = NA)
143     ,panel.background = theme_rect(fill='white', colour='black'))+ geom_smooth(
        method="loess",lwd=1)
144 Forhold = c(forholdRus[,3],forholdNor[,3]);frame = data.frame(Forhold,Land)
145 p=qplot(rep(1960:2008,2),Forhold, data = frame,geom = c("point"),colour=Land,
146 xlab="Periode")
147 #,ylab=expression(frac(E(hat(c)),Sd(hat(c))))

```

8.2. VEDLEGG B - PROGRAMKODE

```
148 p + scale_colour_manual(
149   labels = c("Russland", "Norge"),
150   values = colors()[c(592,134)] + opts(axis.title.y=theme_text(angle=90)
151   ,legend.background = theme_rect(fill = colors()[1], colour = NA)
152   ,panel.background = theme_rect(fill='white', colour='black')) + geom_smooth(
153     method="loess", lwd=1)
154 #Figur 5.9
155 SkjevNor = round((meanThetaN - t(mxThetaN))/t(mxThetaN), 5); SkjevRus = round((
156   meanThetaR - t(mxThetaR))/t(mxThetaR), 5)
157 Skjevhet = c(SkjevRus[,1], SkjevNor[,1]); frame = data.frame(Skjevhet, Land)
158 p = qplot(rep(1960:2008, 2), Skjevhet, data = frame, geom = c("point"), colour=Land,
159   xlab="Periode", ylim=c(-0.26, 0.26))
160 #, ylab=expression(frac(E(hat(a))-a[0], a[0]))
161 p + scale_colour_manual(
162   labels = c("Russland", "Norge"),
163   values = colors()[c(592,134)] + opts(axis.title.y=theme_text(angle=90)
164   ,legend.background = theme_rect(fill = colors()[1], colour = NA)
165   ,panel.background = theme_rect(fill='white', colour='black')) + geom_smooth(
166     method="auto", lwd=1)
167 Skjevhet = c(SkjevRus[,2], SkjevNor[,2]); frame = data.frame(Skjevhet, Land)
168 p = qplot(rep(1960:2008, 2), Skjevhet, data = frame, geom = c("point"), colour=Land,
169   xlab="Periode", ylim=c(-0.03, 0.03))
170 #, ylab=expression(frac(E(hat(b))-b[0], b[0]))
171 p + scale_colour_manual(
172   labels = c("Russland", "Norge"),
173   values = colors()[c(592,134)] + opts(axis.title.y=theme_text(angle=90)
174   ,legend.background = theme_rect(fill = colors()[1], colour = NA)
175   ,panel.background = theme_rect(fill='white', colour='black')) + geom_smooth(
176     method="loess", lwd=1)
177 Skjevhet = c(SkjevRus[,3], SkjevNor[,3]); frame = data.frame(Skjevhet, Land)
178 p = qplot(rep(1960:2008, 2), Skjevhet, data = frame, geom = c("point"), colour=Land,
179   xlab="Periode", ylim=c(-0.006, 0.006))
180 #, ylab=expression(frac(E(hat(c))-c[0], c[0]))
181 p + scale_colour_manual(
182   labels = c("Russland", "Norge"),
183   values = colors()[c(592,134)] + opts(axis.title.y=theme_text(angle=90)
184   ,legend.background = theme_rect(fill = colors()[1], colour = NA)
185   ,panel.background = theme_rect(fill='white', colour='black')) + geom_smooth(
186     method="loess", lwd=1)
187 #Engangspremie
188 onetimeprem = function(q, s, l_r, d)
189 {
190   l_e = length(q); kp = lifetab(q, l_e)$kp
191   l = matrix(0, l_e+1, l_e+1); l[row(l)+col(l)>l_r+1]=1
192   ll=0:l_e; M = s*(d**ll)*kp*1; pi = apply(M, 2, sum); list(pi=pi)}
193 lifetab = function(q, K){
194   l_e = length(q); p = c(1-q, rep(0, l_e+1)); kp = matrix(1, K, l_e+1)
195   for (l in 0:l_e+1){
196     kp[,l] = cumprod(p[1:K+(l-1)])
197   }
198   kp = rbind(rep(1, l_e+1), kp); list(kp=kp)}
199 x=65
200 Pi0.rus = matrix(0, x, year); Pi.rus = matrix(0, x, year); Pi0.nor = matrix(0, x, year)
201 )
202 Pi.nor = matrix(0, x, year); sysRus = matrix(0, x, year); sysNor = matrix(0, x, year)
203 N = 10**6; l_r=65; r=0.03; d=1/(1+r); s=150000
204 for (w in 1:year){
205   Pi.rus[,w] = onetimeprem(mxRus[,w], s, l_r, d)$pi[c(11:x)]
206   Pi0.rus[,w] = onetimeprem(mxRus0[,w], s, l_r, d)$pi[c(11:x)]
207   Pi.nor[,w] = onetimeprem(mxNor[,w], s, l_r, d)$pi[c(11:x)]
208   Pi0.nor[,w] = onetimeprem(mxNor0[,w], s, l_r, d)$pi[c(11:x)]
209   sysRus[,w] = abs(Pi0.rus[,w] - Pi.rus[,w])
210 }
```

8.2. VEDLEGG B - PROGRAMKODE

```
205 sysNor[,w] =abs(Pi0.nor[,w]-Pi.nor[,w])
206 }
207 Nor.Prosent = (sysNor/Pi.nor)*100;Rus.Prosent = (sysRus/Pi.rus)*100
208 meanNOR = apply(Nor.Prosent,2,mean);meanRUS = apply(Rus.Prosent,2,mean)
209 meanSysNor = apply(sysNor,2,mean);meanSysRus = apply(sysRus,2,mean)
210 Sysframe = c(meanNOR,meanRUS)
211 #Figur 5.19b)
212 Land = c(rep("1",49),rep("2",49));frame = data.frame(Sysframe,Land)
213 p=qplot(rep(1960:2008,2),Sysframe, data = frame,geom = c("point","line"),
214         colour=Land,
215 xlab="Periode",ylab="%",main="Relativ systematisk feil i prosent")
216 p + scale_colour_manual(
217     labels = c("Norge","Russland"),
218     values = colors()[c(592,134)]+ opts(axis.title.y=theme_text(angle=0)
219     ,legend.background = theme_rect(fill = colors()[1], colour = NA)
220     ,panel.background = theme_rect(fill='white', colour='black'),legend.text =
221     theme_text(size=14)
222     ,legend.title = theme_text(size = 11, hjust = 15, vjust = 15))
223 #Figur 5.19a)
224 Sysframe = c(meanSysNor,meanSysRus);frame = data.frame(Sysframe,Land)
225 p=qplot(rep(1960:2008,2),Sysframe, data = frame,geom = c("point","line"),
226         colour=Land,
227 xlab="Periode",ylab="Kroner",main="Gjennomsnittelig systematisk feil i kroner"
228 )
229 p + scale_colour_manual(
230     labels = c("Norge","Russland"),
231     values = colors()[c(592,134)]+ opts(axis.title.y=theme_text(angle=90),legend.
232     background = theme_rect(fill = colors()[1], colour = NA)
233     ,panel.background = theme_rect(fill='white', colour='black'),legend.text =
234     theme_text(size=14)
235     ,legend.title = theme_text(size = 11, hjust = 15, vjust = 15))
236 #Systematisk feil i prosent av RKF
237 NN = c(10000,100000,500000,1000000,10000000,1000000000000000);nn=6;x1=1
238 ProsentNor = array(0,dim=c(year,x-x1+1,nn));ProsentRus = array(0,dim=c(year,x-
239 x1+1,nn))
240 for (w in 1:year){
241   for (h in 1:(x-x1+1)){
242     if (sysNor[h,w] < 500) {sysNor[h,w] = 500} else sysNor[h,w] = sysNor[h,w]
243     if (sysRus[h,w] < 500) {sysRus[h,w] = 500} else sysRus[h,w] = sysRus[h,w]
244   }
245   for (k in 1:nn){
246     w11 = Ruspop[,w]/sum(Ruspop[,w]);N = NN[k];n11 = round(w11*N)
247     w12 = Norpop[,w]/sum(Norpop[,w]);N = NN[k];n12 = round(w12*N)
248     Z2 = matrix(rbinom(mb*(12-11+1),n12,mxNor[,w]),12-11+1,mb)
249     Z1 = matrix(rbinom(mb*(12-11+1),n12,mxRus[,w]),12-11+1,mb)
250     Pihat.nor = matrix(0,x-x1+1,mb);Pihat.rus = matrix(0,x-x1+1,mb)
251     for (t in 1:mb){
252       Theta = gompmafit1(l1=l1,n=n11,z=Z1[,t])$theta
253       qTheta = 1-exp(-Theta[1]-Theta[2]*exp(Theta[3]*l1))
254       Pihat.rus[,t] = onetimeprem(qTheta,s,l_r,d)$pi[x1:x]
255       Theta = gompmafit1(l1=l1,n=n12,z=Z2[,t])$theta
256       qTheta = 1-exp(-Theta[1]-Theta[2]*exp(Theta[3]*l1))
257       Pihat.nor[,t] = onetimeprem(qTheta,s,l_r,d)$pi[x1:x]
258     }
259     TOTAL1 = sqrt(apply((sysRus[,w]**2+(Pi0.rus[,w]-Pihat.rus)**2),1,mean))
260     TOTAL2 = sqrt(apply((sysNor[,w]**2+(Pi0.nor[,w]-Pihat.nor)**2),1,mean))
261     ProsentNor[w,,k] = sysNor[,w]/TOTAL2
262     ProsentRus[w,,k] = sysRus[,w]/TOTAL1
263   }
264 }
265 ProsentMeanNor = matrix(0,year,nn);ProsentMeanRus = matrix(0,year,nn)
266 for (i in 1:6){
267   ProsentMeanNor[,i] = apply(ProsentNor[, ,i],1,mean)
268   ProsentMeanRus[,i] = apply(ProsentRus[, ,i],1,mean)
```

8.2. VEDLEGG B - PROGRAMKODE

```
261 }
262 #Figur 5.20a)
263 systematiskframe = c(100*ProsentMeanNor)
264 Eksponeringsstørrelse = c(rep("1",49),rep("2",49),rep("3",49),rep("4",49),rep(
  "5",49),rep("6",49))
265 frame = data.frame(systematiskframe,Eksponeringsstørrelse)
266 p=qplot(rep(1960:2008,6),systematiskframe, data = frame,geom = c("line"),
  colour=Eksponeringsstørrelse,
267 xlab="Periode",ylab="%",
268 ,main="Systematisk feil i prosent, Norge")
269 p + scale_colour_manual(name = "Eksponeringsstørrelse",
270                          labels = c("N = 10,000", "N = 100,000", "N = 500,000", "
  N = 1,000,000","N = 10,000,000","N > 50,000,000"),
271 values = c(1:6)) + guides(col = guide_legend(reverse = TRUE))+ opts(axis.title
  .y=theme_text(angle=0)
272 ,legend.background = theme_rect(fill = colors()[1], colour = NA)
273 ,panel.background = theme_rect(fill='white', colour='black'))
274 #Figur 5.20b)
275 systematiskframe = c(100*ProsentMeanRus);frame = data.frame(systematiskframe,
  Eksponeringsstørrelse)
276 p=qplot(rep(1960:2008,6),systematiskframe, data = frame,geom = c("line"),
  colour=Eksponeringsstørrelse,
277 xlab="Periode",ylim=c(-0.005,100),ylab = "%")
278 ,main="Systematisk feil i prosent, Russland")
279 p + scale_colour_manual(name = "Eksponeringsstørrelse",
280                          labels = c("N = 10,000", "N = 100,000", "N = 500,000", "
  N = 1,000,000","N = 10,000,000","N > 50,000,000"),
281 values = c(1:6)) + guides(col = guide_legend(reverse = TRUE))+ opts(axis.title
  .y=theme_text(angle=0)
282 ,legend.background = theme_rect(fill = colors()[1], colour = NA)
283 ,panel.background = theme_rect(fill='white', colour='black'))
284 ##BETA-BINOMISK
285 JapMx <- hmd.mx("JPN", username="lars.oesthassel@gmail.com", password="XXX",
  Japan")
286 UsaMx <- hmd.mx("USA", username="lars.oesthassel@gmail.com", password="XXX",
  Usa")
287 SweMx <- hmd.mx("SWE", username="lars.oesthassel@gmail.com", password="XXX",
  Sweden")
288 ItaMx <- hmd.mx("ITA", username="lars.oesthassel@gmail.com", password="XXX",
  Italy")
289 l1 = 1:100
290 Y1=NorMx$pop$male[l1,114:164]
291 rate1=NorMx$rate$male[l1,114:164]
292 Z1=Y1*rate1
293 Y2=RusMx$pop$male[l1,1:51]
294 rate2=RusMx$rate$male[l1,1:51]
295 Z2=Y2*rate2
296 Y3=JapMx$pop$male[l1,13:63]
297 rate3=JapMx$rate$male[l1,13:63]
298 Z3=Y3*rate3
299 Y4=UsaMx$pop$male[l1,28:78]
300 rate4=UsaMx$rate$male[l1,28:78]
301 Z4=Y4*rate4
302 Y5=SweMx$pop$male[l1,211:261]
303 rate5=SweMx$rate$male[l1,211:261]
304 Z5=Y5*rate5
305 Y6=ItaMx$pop$male[l1,88:138]
306 rate6=ItaMx$rate$male[l1,88:138]
307 Z6=Y6*rate6
308 func=function(s,l1,n,z)
309 {
310 t=exp(s)
311 xi = (1-pmax(exp(-t[1]-t[2]*(exp(t[3]*l1))),0.0000000000000001))
```

8.2. VEDLEGG B - PROGRAMKODE

```
312 ax = t[4]*xi
313 bx = t[4]*(1-xi)
314 -sum(lgamma(ax+z)+lgamma(n-z+bx)
315 +lgamma(bx+ax)-lgamma(ax+bx+n)-lgamma(ax)-lgamma(bx))
316 }
317 Beta=function(l1,n,z,gam){
318 s=c(-8,-9,-2,log(gam))
319 #o=optim(s,func,"BFGS",l1=l1,n=n,z=z)
320 o=optim(s,func,l1=l1,n=n,z=z)
321 }
322 year = length(Y1[,])
323 l1 = 1
324 l2 = length(Y1[,1])
325 l1 = 1:l2
326 QQnor = matrix(0,4,year)
327 QQrus = matrix(0,4,year)
328 QQjap = matrix(0,4,year)
329 QQusa = matrix(0,4,year)
330 QQswe = matrix(0,4,year)
331 QQita = matrix(0,4,year)
332 for (i in 1:year){
333   QQnor[,i] = Beta(l1,Y1[(l1+l2*(i-1)):(l2+l2*(i-1))],Z1[(l1+l2*(i-1)):(l2*i)
334     ],700)$par
335   QQrus[,i] = Beta(l1,Y2[(l1+l2*(i-1)):(l2+l2*(i-1))],Z2[(l1+l2*(i-1)):(l2*i)
336     ],4000)$par
337   QQjap[,i] = Beta(l1,Y3[(l1+l2*(i-1)):(l2+l2*(i-1))],Z3[(l1+l2*(i-1)):(l2*i)
338     ],500)$par
339   QQusa[,i] = Beta(l1,Y4[(l1+l2*(i-1)):(l2+l2*(i-1))],Z4[(l1+l2*(i-1)):(l2*i)
340     ],4000)$par
341   QQswe[,i] = Beta(l1,Y5[(l1+l2*(i-1)):(l2+l2*(i-1))],Z5[(l1+l2*(i-1)):(l2*i)
342     ],4000)$par
343   QQita[,i] = Beta(l1,Y6[(l1+l2*(i-1)):(l2+l2*(i-1))],Z6[(l1+l2*(i-1)):(l2*i)
344     ],4000)$par
345 }
346 Gammaframe = c(exp(QQnor[4,]),exp(QQrus[4,]),exp(QQjap[4,]),exp(QQusa[4,]),exp(
347   QQswe[4,]),exp(QQita[4,]))
348 Land = c(rep("Norge",51),rep("Russland",51),rep("Japan",51),rep("Usa",51),rep(
349   "Sverige",51),rep("Italia",51))
350 frame = data.frame(Gammaframe,Land)
351 library(ggplot2)
352 p=qplot(rep(1959:2009,6),Gammaframe, data = frame,geom = c("point","line"),
353   colour=Land,xlab="Periode",ylab=expression(gamma))
354 p + opts(axis.title.y=theme_text(angle=0)
355   ,panel.background = theme_rect(fill='white', colour='black'))
```

ManyYears.R

8.2.4 Kapittel 6

6.1-6.3 - LC

```
1 library(demography);setwd("C:/Users/Lars/Desktop/ilc");attach("ilc.rdata");
2   attach("cmi.rdata");args(lca.rh)
3 NorMx <- hmd.mx("NOR", username="lars.oesthassel@gmail.com", password="XXX",
4   Norway")
5 #Function
6 Demo = function(x,y){
7   demogdata(data=NorMx$rate$male[1:100,x:y],pop=NorMx$pop$male[1:100,x:y],name="
8     male",
```

8.2. VEDLEGG B - PROGRAMKODE

```
6 ages=NorMx$age[1:100],years=NorMx$year[x:y],type="mortality",label="Norge",
  lambda=1)
7 }
8 kk1945=lca.rh(Demo(100,164),mod="lc",interpolate=T,verbose=F)
9 ax1945 = kk1945$ax;bx1945 = kk1945$bx;kt1945 = kk1945$kt
10 kk1960=lca.rh(Demo(114,164),mod="lc",interpolate=T,verbose=F)
11 ax1960 = kk1960$ax;bx1960 = kk1960$bx;kt1960 = kk1960$kt
12 kk1970=lca.rh(Demo(124,164),mod="lc",interpolate=T,verbose=F)
13 ax1970 = kk1970$ax;bx1970 = kk1970$bx;kt1970 = kk1970$kt
14 kk1980=lca.rh(Demo(134,164),mod="lc",interpolate=T,verbose=F)
15 ax1980 = kk1980$ax;bx1980 = kk1980$bx;kt1980 = kk1980$kt
16 #Figur 6.2
17 par(mfrow=c(1,3))
18 ###1980-2010###
19 T = length(kt1980);kt1 = kt1980[1];ktend = kt1980[T]
20 drift = (ktend-kt1)/(T-1);sigma = (1/(T-1))*sum((kt1980[-1]-kt1980[-T]-drift)
  ^2)
21 mb = 1000;y = 60;Ty = T + y +1;KT = matrix(ktend,mb,y+1)
22 for (i in 1:y){
23   KT[,i+1] = KT[,i] + drift + rnorm(mb,0,sqrt(sigma))
24 }
25 ktpredicted = apply(KT,2,mean);s = c(0.025,0.975);pick = mb*s;KTsort = apply(
  KT,2,sort)
26 Ktsort1980 = KTsort[pick,];fullkt1980 = c(kt1980,ktpredicted)
27 mx1980 = matrix(0,100,Ty);mxup1980 = matrix(0,100,y+1);mxlow1980 = matrix
  (0,100,y+1)
28 for (i in 1:100){
29   mx1980[i,] = exp(ax1980[i]+bx1980[i]*fullkt1980)
30   mxup1980[i,] = exp(ax1980[i]+bx1980[i]*Ktsort1980[1,])
31   mxlow1980[i,] = exp(ax1980[i]+bx1980[i]*Ktsort1980[2,])
32 }
33 forventet1980 = 0.5+apply(apply(1-mx1980,2,cumprod),2,sum)
34 forventetlow1980 = 0.5+apply(apply(1-mxlow1980,2,cumprod),2,sum)
35 forventetup1980 = 0.5+apply(apply(1-mxup1980,2,cumprod),2,sum)
36 qx = NorMx$rate$male[1:100,134:164];forventetSann = 0.5+apply(apply(1-qx,2,
  cumprod),2,sum)
37 plot(ts(forventetSann,start=1979),main="1980-2070",xlab="Periode",ylab="
  Forventet levealder",ylim=c(67,86),xlim=c(1979,2060),lwd=2)
38 legend(coord("UL"), legend = c("Observerv", "Tilpasset", "Predikert"), lty = c
  (1,
39   3, 2), col = c(1, "red", "blue"))
40 fcast = cbind(c(2010:2070),forventetlow1980,forventetup1980)
41 filled.gap(fcast,edges=F,app=T)
42 lines(ts(forventet1980[32:92],start=2010),lwd=2,col="blue",lty=3)
43 lines(ts(forventet1980[1:31],start=1979),lwd=2,col="red",lty=3)
44 text(2045, 79, "Predikert");text(1990, 77, "Tilpasset")
45 ###1960-2010###
46 T = length(kt1960);kt1 = kt1960[1];ktend = kt1960[T];drift = (ktend-kt1)/(T-1)
47 sigma = (1/(T-1))*sum((kt1960[-1]-kt1960[-T]-drift)^2);mb = 1000;y = 60
48 Ty = T + y +1;KT = matrix(ktend,mb,y+1)
49 for (i in 1:y){
50   KT[,i+1] = KT[,i] + drift + rnorm(mb,0,sqrt(sigma))
51 }
52 ktpredicted = apply(KT,2,mean);s = c(0.025,0.975);pick = mb*s;KTsort = apply(
  KT,2,sort)
53 Ktsort1960 = KTsort[pick,];fullkt1960 = c(kt1960,ktpredicted);mx1960 = matrix
  (0,100,Ty)
54 mxup1960 = matrix(0,100,y+1);mxlow1960 = matrix(0,100,y+1)
55 for (i in 1:100){
56   mx1960[i,] = exp(ax1960[i]+bx1960[i]*fullkt1960)
57   mxup1960[i,] = exp(ax1960[i]+bx1960[i]*Ktsort1960[1,])
58   mxlow1960[i,] = exp(ax1960[i]+bx1960[i]*Ktsort1960[2,])
59 }
```


8.2. VEDLEGG B - PROGRAMKODE

```
60 forventet1960 = 0.5+apply(apply(1-mx1960,2,cumprod),2,sum)
61 forventetlow1960 = 0.5+apply(apply(1-mxlow1960,2,cumprod),2,sum)
62 forventetup1960 = 0.5+apply(apply(1-mxup1960,2,cumprod),2,sum)
63 qx = NorMx$rate$male[1:100,114:164];forventetSann = 0.5+apply(apply(1-qx,2,
  cumprod),2,sum)
64 plot(ts(forventetSann,start=1959),main="1960-2070",xlab="Periode",ylab="
  Forventet levealder",ylim=c(67,86),xlim=c(1959,2060),lwd=2)
65 legend(coord("UL"), legend = c("Observert", "Tilpasset", "Predikert"), lty = c
  (1,
    3, 2), col = c(1, "red", "blue"))
66 fcast = cbind(c(2010:2070),forventetlow1960,forventetup1960)
67 filled.gap(fcast,edges=F,app=T)
69 lines(ts(forventet1960[52:113],start=2010),lwd=2,col="blue",lty=3)
70 lines(ts(forventet1960[1:51],start=1959),lwd=2,col="red",lty=3)
71 text(2040, 78, "Predikert");text(1975, 75, "Tilpasset")
72 ###1945-2010###
73 T = length(kt1945);kt1 = kt1945[1];ktend = kt1945[T]
74 drift = (ktend-kt1)/(T-1);sigma = (1/(T-1))*sum((kt1945[-1]-kt1945[-T]-drift)
  ^2);Ty = T + y + 1
75 KT = matrix(ktend,mb,y+1)
76 for (i in 1:y){
77   KT[,i+1] = KT[,i] + drift + rnorm(mb,0,sqrt(sigma))
78 }
79 ktpredicted = apply(KT,2,mean);s = c(0.025,0.975);pick = mb*s;Ktsort = apply(
  KT,2,sort)
80 Ktsort1945 = Ktsort[pick,];fullkt1945 = c(kt1945,ktpredicted)
81 mx1945 = matrix(0,100,Ty);mxup1945 = matrix(0,100,y+1);mxlow1945 = matrix
  (0,100,y+1)
82 for (i in 1:100){
83   mx1945[i,] = exp(ax1945[i]+bx1945[i]*fullkt1945)
84   mxup1945[i,] = exp(ax1945[i]+bx1945[i]*Ktsort1945[1,])
85   mxlow1945[i,] = exp(ax1945[i]+bx1945[i]*Ktsort1945[2,])
86 }
87 forventet1945 = 0.5+apply(apply(1-mx1945,2,cumprod),2,sum)
88 forventetlow1945 = 0.5+apply(apply(1-mxlow1945,2,cumprod),2,sum)
89 forventetup1945 = 0.5+apply(apply(1-mxup1945,2,cumprod),2,sum)
90 qx = NorMx$rate$male[1:100,100:164];forventetSann = 0.5+apply(apply(1-qx,2,
  cumprod),2,sum)
91 plot(ts(forventetSann,start=1945),main="1945-2070",xlab="Periode",ylab="
  Forventet levealder",ylim=c(67,86),xlim=c(1945,2060),lwd=2)
92 legend(coord("UL"), legend = c("Observert", "Tilpasset", "Predikert"), lty = c
  (1,
    3, 2), col = c(1, "red", "blue"))
93 fcast = cbind(c(2010:2070),forventetlow1945,forventetup1945)
94 filled.gap(fcast,edges=F,app=T)
96 lines(ts(forventet1945[66:128],start=2010),lwd=2,col="blue",lty=3)
97 lines(ts(forventet1945[1:65],start=1945),lwd=2,col="red",lty=3)
98 text(2040, 78, "Predikert");text(1970, 75, "Tilpasset")
99 #Figur 6.1
100 trueframe = c(fullkt1945,fullkt1960,fullkt1980,Ktsort1945[1,],Ktsort1945[2,],
101 Ktsort1960[1,],Ktsort1960[2,],Ktsort1980[1,],Ktsort1980[2,])
102 Modell = c(rep("1",126),rep("2",112),rep("3",92),rep("1",122),rep("2",122),rep
  ("3",122))
103 frame2 = data.frame(trueframe,Modell)
104 p=qplot(c(1945:2070,1959:2070,1979:2070,rep(2010:2070,6)),trueframe, data =
  frame2,geom = c("path","line"),colour=Modell,
105 xlab="Periode",ylab=expression(k[k]))
106 p + annotate("text", label = c("Predikert","Tilpasset"), x = c(2045,1970),y =
  c(0,-50), size = 6) + scale_colour_manual(
107   labels = c("LC1945","LC1960","LC1980"),
108   values = colors()[c(134,24,254)]+ opts(axis.title.y=theme_text(angle=0)
109   ,legend.background = theme_rect(fill = colors()[1], colour = NA)
110   ,panel.background = theme_rect(fill='white', colour='black'))
```

8.2. VEDLEGG B - PROGRAMKODE

```
111 |
112 | #Tabell 6.1
113 | forventet1980[c(32,42,52,62,72,82)];forventet1960[c(52,62,72,82,92,102)];
    |   forventet1945[c(66,76,86,96,106,116)]
114 | c(forventet1980[32],forventet1980[52],forventetlow1980[21],forventetup1980
    |   [21],forventet1980[92],forventetlow1980[61],forventetup1980[61])
115 | c(forventet1960[52],forventet1960[72],forventetlow1960[21],forventetup1960
    |   [21],forventet1960[112],forventetlow1960[61],forventetup1960[61])
116 | c(forventet1945[66],forventet1945[86],forventetlow1945[21],forventetup1945
    |   [21],forventet1945[126],forventetlow1945[61],forventetup1945[61])
117 | S = array(0,dim=c(length(ax1980),length(32:(32+y)),10))
118 | gl = NorMx$rate$male[1:100,163]
119 | S[,1] = mx1980[-c(1:31)];S[,2] = mxlow1980;S[,3] = mxup1980;S[,4] =
    |   mx1960[-c(1:51)]
120 | S[,5] = mxlow1960;S[,6] = mxup1960;S[,7] = mx1945[-c(1:65)];S[,8] =
    |   mxlow1945;S[,9] = mxup1945
121 | for (i in 1:61){
122 |   S[,i,10] = gl
123 | }
124 | 0.5+apply(apply(1-S[,4],2,cumprod),2,sum);0.5+apply(apply(1-S[,7],2,cumprod)
    |   ,2,sum)
125 | 0.5+apply(apply(1-S[,1],2,cumprod),2,sum);s = 150000;r = 0.03;d = 1/(1+r);lr
    |   = 65
126 | l1 = 40;l2 = length(ax1980);pi.ALL = matrix(0,l2-l1+1,10)
127 | for (i in 1:10){
128 |   for (l in l1:l2){
129 |     k = 1:l2
130 |     q = cumprod(diag(1-S[k,,i]))
131 |     K = max(lr-l,0)
132 |     pi.ALL[l-l1+1,i] = s*sum(q[((K+1):(l2-l+1))]*d^(K:(l2-l)))
133 |   }
134 | library(ggplot2)
135 | Dynamisk = c(pi.ALL[1:40,c(1,4,7,10)])
136 | Modell = c(rep("1",40),rep("2",40),rep("3",40),rep("4",40))
137 | frame = data.frame(Dynamisk,Modell)
138 | p=qplot(rep(40:79,4),Dynamisk, data = frame,geom = c("line"),colour=Modell,
139 |   xlab="Alder",ylab="Engangpremie i kroner")
140 | #Figur 6.4
141 | p + scale_colour_manual(
142 |   labels = c("LC1980","LC1960","LC1945","Statisk GM"),
143 |   values = colors()[c(254,24,134,461)] + opts(axis.title.y=theme_text(angle=90)
144 | ,legend.background = theme_rect(fill = colors()[1], colour = NA)
145 | ,panel.background = theme_rect(fill='white', colour='black'))
146 | l1 = 40;l2 = 100;kpk1980 = apply(1-mx1980[l1:l2,-c(1:31)],2,cumprod)
147 | kpk1960 = apply(1-mx1960[l1:l2,-c(1:51)],2,cumprod);kpk1945 = apply(1-mx1945[
    |   l1:l2,-c(1:65)],2,cumprod)
148 | e1980 = numeric(0);e1960 = numeric(0);e1945 = numeric(0);N = 10000
149 | for (j in 25:61){
150 |   e1980[j-24] = kpk1980[j,j]*N
151 |   e1960[j-24] = kpk1960[j,j]*N
152 |   e1945[j-24] = kpk1945[j,j]*N
153 | }
154 | e = cbind(e1980,e1960,e1945);r = 1.03;NN = 10^9;s = 150000/NN;VPre = rep(25,3)
    |   ;VPre[1] = pi.ALL[1,10]*N/NN;VPre[2:25] = VPre[1]*1.03^(1:24)
155 | V1980 = rep(VPre[25],length(25:61));V1960 = rep(VPre[25],length(25:61));V1945
    |   = rep(VPre[25],length(25:61))
156 | for (i in 1:36){
157 |   if (V1980[i]>0)
158 |     V1980[i+1] = V1980[i]*r - e1980[i]*s
159 |   else
160 |     V1980[i+1] = V1980[i] - e1980[i]*s
161 |   if (V1960[i]>0)
162 |     V1960[i+1] = V1960[i]*r - e1960[i]*s
```

8.2. VEDLEGG B - PROGRAMKODE

```

163   else
164     V1960[i+1] = V1960[i] - e1960[i]*s
165   if (V1945[i]>0)
166     V1945[i+1] = V1945[i]*r - e1945[i]*s
167   else
168     V1945[i+1] = V1945[i] - e1945[i]*s
169 }
170 Vakk1980 = c(VPre,V1980[2:36]);Vakk1980[60] = Vakk1980[60]+s*e1980[37];
    Vakk1960 = c(VPre,V1960[2:36])
171 Vakk1960[60] = Vakk1960[60]+s*e1960[37];Vakk1945 = c(VPre,V1945[2:36]);
    Vakk1945[60] = Vakk1945[60]+s*e1945[37]
172 Dynamisk = c(Vakk1980,Vakk1960,Vakk1945);Modell = c(rep("1",60),rep("2",60),
    rep("3",60));frame = data.frame(Dynamisk,Modell)
173 p=qplot(rep(2010:2069,3),Dynamisk, data = frame,geom = c("line"),colour=Modell
    ,
174 xlab="Periode",ylab="Akkumulert porteføljeverdi")
175 #Figur 6.5
176 p + annotate("text", label = c("Utbetalingsperiode","Investeringsperiode")
177 , x = c(2055,2022),y = c(10,6), size = 5) + scale_colour_manual(
178     labels = c("LC1980","LC1960","LC1945"),
179 values = colors()[c(254,24,134)]) + opts(axis.title.y=theme_text(angle=90)
180 ,legend.background = theme_rect(fill = colors()[1], colour = NA)
181 ,panel.background = theme_rect(fill='white', colour='black'))
182 #Tabell 6.2
183 V1980[c(1,11,21,31,37)];V1960[c(1,11,21,31,37)];V1945[c(1,11,21,31,37)]
184 ###Backtesting###
185 kk1945=lca.rh(Demo(100,154),mod="lc",interpolate=T,verbose=F)
186 ax1945 = kk1945$ax;bx1945 = kk1945$bx;kt1945 = kk1945$kt
187 kk1960=lca.rh(Demo(114,154),mod="lc",interpolate=T,verbose=F)
188 ax1960 = kk1960$ax;bx1960 = kk1960$bx;kt1960 = kk1960$kt
189 kk1980=lca.rh(Demo(134,154),mod="lc",interpolate=T,verbose=F)
190 ax1980 = kk1980$ax;bx1980 = kk1980$bx;kt1980 = kk1980$kt
191 kk1990=lca.rh(Demo(145,154),mod="lc",interpolate=T,verbose=F)
192 ax1990 = kk1990$ax;bx1990 = kk1990$bx;kt1990 = kk1990$kt
193 T1990 = length(kt1990);T1980 = length(kt1980);T1960 = length(kt1960);T1945 =
    length(kt1945)
194 kt0 = kt1990[1];kt1 = kt1980[1];kt2 = kt1960[1];kt3 = kt1945[1]
195 ktend1990 = kt1990[T1990];ktend1980 = kt1980[T1980];ktend1960 = kt1960[T1960];
    ktend1945 = kt1945[T1945]
196 drift1990 = (ktend1990-kt0)/(T1990-1);drift1980 = (ktend1980-kt1)/(T1980-1)
197 drift1960 = (ktend1960-kt2)/(T1960-1);drift1945 = (ktend1945-kt3)/(T1945-1)
198 sigma1990 = (1/(T1990-1))*sum((kt1990[-1]-kt1990[-T1990]-drift1990)^2)
199 sigma1980 = (1/(T1980-1))*sum((kt1980[-1]-kt1980[-T1980]-drift1980)^2)
200 sigma1960 = (1/(T1960-1))*sum((kt1960[-1]-kt1960[-T1960]-drift1960)^2)
201 sigma1945 = (1/(T1945-1))*sum((kt1945[-1]-kt1945[-T1945]-drift1945)^2)
202 mb = 10000;y = 10;Ty1990 = T1990 + y +1;Ty1980 = T1980 + y +1
203 Ty1960 = T1960 + y +1;Ty1945 = T1945 + y +1;KT1990 = matrix(ktend1990,mb,y+1)
204 KT1980 = matrix(ktend1980,mb,y+1);KT1960 = matrix(ktend1960,mb,y+1);KT1945 =
    matrix(ktend1945,mb,y+1)
205 for (i in 1:y){
206   KT1990[,i+1] = KT1990[,i] + drift1990 + rnorm(mb,0,sqrt(sigma1990))
207   KT1980[,i+1] = KT1980[,i] + drift1980 + rnorm(mb,0,sqrt(sigma1980))
208   KT1960[,i+1] = KT1960[,i] + drift1960 + rnorm(mb,0,sqrt(sigma1960))
209   KT1945[,i+1] = KT1945[,i] + drift1945 + rnorm(mb,0,sqrt(sigma1945))
210 }
211 ktpred1990 = apply(KT1990,2,mean);ktpred1980 = apply(KT1980,2,mean)
212 ktpred1960 = apply(KT1960,2,mean);ktpred1945 = apply(KT1945,2,mean)
213 s = c(0.025,0.975);pick = mb*s;KTsort = apply(KT1990,2,sort);Ktsort1990 =
    KTsort[pick,]
214 fullkt1990 = c(kt1990,ktpred1990);KTsort = apply(KT1980,2,sort);Ktsort1980 =
    KTsort[pick,]
215 fullkt1980 = c(kt1980,ktpred1980);KTsort = apply(KT1960,2,sort);Ktsort1960 =
    KTsort[pick,]

```

8.2. VEDLEGG B - PROGRAMKODE

```
216 fullkt1960 = c(kt1960, ktpred1960); Ktsort = apply(KT1945, 2, sort); Ktsort1945 =
    Ktsort[pick,]
217 fullkt1945 = c(kt1945, ktpred1945); mx1990 = matrix(0, 100, Ty1990); mx1980 =
    matrix(0, 100, Ty1980)
218 mx1960 = matrix(0, 100, Ty1960); mx1945 = matrix(0, 100, Ty1945); mxup1945 = matrix
    (0, 100, y+1); mxlow1945 = matrix(0, 100, y+1)
219 mxup1980 = matrix(0, 100, y+1); mxlow1980 = matrix(0, 100, y+1); mxup1960 = matrix
    (0, 100, y+1); mxlow1960 = matrix(0, 100, y+1)
220 for (i in 1:100){
221   mx1990[i,] = exp(ax1990[i]+bx1990[i]*fullkt1990)
222   mx1980[i,] = exp(ax1980[i]+bx1980[i]*fullkt1980)
223   mxup1980[i,] = exp(ax1980[i]+bx1980[i]*Ktsort1980[1,])
224   mxlow1980[i,] = exp(ax1980[i]+bx1980[i]*Ktsort1980[2,])
225   mx1960[i,] = exp(ax1960[i]+bx1960[i]*fullkt1960)
226   mxup1960[i,] = exp(ax1960[i]+bx1960[i]*Ktsort1960[1,])
227   mxlow1960[i,] = exp(ax1960[i]+bx1960[i]*Ktsort1960[2,])
228   mx1945[i,] = exp(ax1945[i]+bx1945[i]*fullkt1945)
229   mxup1945[i,] = exp(ax1945[i]+bx1945[i]*Ktsort1945[1,])
230   mxlow1945[i,] = exp(ax1945[i]+bx1945[i]*Ktsort1945[2,])
231 }
232 SPre = array(0, dim=c(length(ax1980), 11, 10))
233 SPre[, , 1] = mx1980[, (T1980+1):Ty1980]; SPre[, , 2] = mxlow1980; SPre[, , 3] =
    mxup1980; SPre[, , 4] = mx1960[, (T1960+1):Ty1960]
234 SPre[, , 5] = mxlow1960; SPre[, , 6] = mxup1960; SPre[, , 7] = mx1945[, (T1945+1):
    Ty1945]; SPre[, , 8] = mxlow1945
235 SPre[, , 9] = mxup1945; SPre[, , 10] = mx1990[, (T1990+1):Ty1990]; Forventet.Pred =
    matrix(0, 11, 10)
236 Forventet.Fit1990 = ts(0.5 + apply(apply(1-mx1990, 2, cumprod), 2, sum), start =
    1990)
237 Forventet.Fit1980 = ts(0.5 + apply(apply(1-mx1980, 2, cumprod), 2, sum), start =
    1979)
238 Forventet.Fit1960 = ts(0.5 + apply(apply(1-mx1960, 2, cumprod), 2, sum), start =
    1959)
239 Forventet.Fit1945 = ts(0.5 + apply(apply(1-mx1945, 2, cumprod), 2, sum), start =
    1945)
240 for (i in 1:10){
241   Forventet.Pred[, i] = ts(0.5 + apply(apply(1-SPre[, , i], 2, cumprod), 2, sum),
    start=2000)
242 }
243 Forventet.True = ts(0.5+apply(apply(1-NorMx$rate$male[1:100, 144:164], 2, cumprod
    ), 2, sum), start=2000)
244 Dynamisk = c(Forventet.True, Forventet.Fit1990, Forventet.Fit1980[12:32],
    Forventet.Fit1960[32:52],
245 (c(Forventet.Fit1945[46:53], 0.7+Forventet.Fit1945[54:66])))
246 Modell = c(rep("1", 21), rep("2", 21), rep("3", 21), rep("4", 21), rep("5", 21))
247 frame = data.frame(Dynamisk, Modell)
248 #Figur 5.24
249 p=qplot(rep(1990:2010, 5), Dynamisk, data = frame, geom = c("line"), colour=Modell
    ,
250 xlab="Periode", ylab="Forventet Levealder")
251 #Figur 6.3
252 p + annotate("text", label = c("Tilpasset", "Predikert"), x = c(1992, 2003), y =
    c(75, 77.5), size = 5) + scale_colour_manual(
253   labels = c("Sann Verdi", "LC1990", "LC1980", "LC1960", "
    LC1945"),
254 values = colors()[c(450, 73, 254, 24, 134)]) + opts(axis.title.y=theme_text(angle
    =90)
255 , legend.background = theme_rect(fill = colors()[1], colour = NA)
256 , panel.background = theme_rect(fill='white', colour='black'))
```

LCNorge.R

Bibliografi

- [1] Erik Bølviken. Notes and curriculum in stk 4520 and 4540, university of oslo. *To be published, Cambridge University Press.*
- [2] KO Bowman and LR Shenton. Parameter estimation for the beta distribution. *Journal of statistical computation and simulation*, 43(3-4):217–228, 1992.
- [3] Natacha Brouhns, Michel Denuit, and Jeroen K Vermunt. A poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31(3):373–393, 2002.
- [4] Zoltan Butt and Steve Haberman. Ilc: A collection of r functions for fitting a class of lee-carter mortality models using iterative fitting algorithms. *Actuarial Research Paper*, (190), 2009.
- [5] Jay L Devore and Kenneth N Berk. *Modern mathematical statistics with applications*. Brooks/Cole, 2007.
- [6] Avinash K Dixit and Joseph E Stiglitz. Monopolistic competition and optimum product diversity. *The American Economic Review*, 67(3):297–308, 1977.
- [7] Kevin Dowd, Andrew JG Cairns, David Blake, Guy D Coughlan, David Epstein, and Marwa Khalaf-Allah. Backtesting stochastic mortality models: An ex post evaluation of multiperiod-ahead density forecasts. *North American Actuarial Journal*, 14(3):281–298, 2010.
- [8] Cheng-Der Fuh. Efficient likelihood estimation in state space models. *The Annals of Statistics*, pages 2026–2068, 2006.
- [9] Leonid A Gavrilov, Natalia S Gavrilova, and Victor N Nosov. Human life span stopped increasing: why? *Gerontology*, 29(3):176–180, 2009.
- [10] Benjamin Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical transactions of the Royal Society of London*, 115:513–583, 1825.
- [11] DA Griffiths. Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. *Biometrics*, pages 637–648, 1973.

BIBLIOGRAFI

- [12] Peter J Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 221–33, 1967.
- [13] Rob J Hyndman and Anne B Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006.
- [14] Nico Keilman. Demography: uncertain population forecasts. *Nature*, 412(6846):490–491, 2001.
- [15] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.
- [16] Ronald Lee. The lee-carter method for forecasting mortality, with various extensions and applications. *North American actuarial journal*, 4(1):80–91, 2000.
- [17] Ronald D Lee and Lawrence R Carter. Modeling and forecasting us mortality. *Journal of the American statistical association*, 87(419):659–671, 1992.
- [18] William Matthew Makeham. On the law of mortality and the construction of annuity tables. *The Assurance Magazine, and Journal of the Institute of Actuaries*, 8(6):301–310, 1860.
- [19] Ermanno Pitacco. Survival models in a dynamic context: a survey. *Insurance: Mathematics and Economics*, 35(2):279–298, 2004.
- [20] Geir Storvik. Bootstrapping–tilleggs litteratur for stk1120. 2005.
- [21] Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, pages 1–25, 1982.
- [22] Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009. ISBN 978-0-387-98140-6. URL <http://had.co.nz/ggplot2/book>.
- [23] John R Wilmoth and Vladimir Shkolnikov. Human mortality database. 2006.
- [24] Rob J Hyndman with contributions from Heather Booth, Leonie Tickle, John Maindonald, Simon Wood, and R Core Team. *demography: Forecasting mortality, fertility, migration and population data*, 2012. URL <http://CRAN.R-project.org/package=demography>. R package version 1.14.
- [25] David Zaridze, Paul Brennan, Jillian Boreham, Alex Boroda, Rostislav Karpov, Alexander Lazarev, Irina Konobeevskaya, Vladimir Igitov, Tatiana Terechova, Paolo Boffetta, et al. Alcohol and cause-specific mortality in russia: a retrospective case-control study of 48 557 adult deaths. *The Lancet*, 373(9682):2201–2214, 2009.